

Clinton wins a great victory on NAFTA

The US Congress, largely because of bipartisan support, has voted to ratify the North American Free Trade Agreement and thereby enhanced the status of US President Bill Clinton in the international arena.

ALL across North America last week, ordinary citizens in the United States, Canada and Mexico joined in a debate about the pros and cons of the North American Free Trade Agreement (NAFTA) that was up for a vote on 17 November by the US House of Representatives. The treaty was strongly backed by President Bill Clinton, who put the prestige of the presidency on the line to win an astounding victory with a vote of 234:200. The US Senate on 20 November voted overwhelmingly (61:38) to ratify the treaty, making Clinton's victory complete.

It is unusual to hear trade and tariffs debated with equal vigour in the bars and board rooms of the United States, but a massive display of journalistic overkill made the treaty the topic of the week. Much of it was egged on by billionaire Ross Perot who on radio and television gave it as his opinion that passage of NAFTA would lead to a "great sucking sound" as American jobs moved south to Mexico, and hyperbolically turned the debate into a class war, pitting anti-NAFTA blue-collar workers and unions as the good guys against the pro-NAFTA elite.

The basic shape of NAFTA has been around since former US president George Bush backed the free-trade agreement late in his administration. But it was Clinton who took it on as a serious job. In so doing he roused the anger of labour unions and other significant elements in the Democratic party who seemed to accept Perot's diagnosis. To win in the House, Clinton had to win the support of a majority of Republicans and will, therefore, have to mend some Democratic fences as he takes on his next great challenge — health-care reform.

But overall it was worth the fight. Passage of the treaty will, over time, be good for the workers and economies of all three North American countries as tariffs on virtually all goods are gradually eliminated during a period of 15 years. As has already been noted (see *Nature* 365, 1; 1993), NAFTA is not as sweeping an agreement as those that created the European Economic Community, which permits the free exchange of people as well as goods. NAFTA is limited to eliminating tariffs that have hampered free trade in North America, but that is no small matter.

Technology and manufacturing companies, agricultural producers and others will now be able to do business with an eye to efficiency, quality and scientific enhancement of the workplace. Automobile manufacturers, for instance, will no longer need to produce cars in Mexico, the United States and

Canada in order to sell in all three markets.

This is not a treaty that will revolutionize economies overnight, but its significance as an indication that the United States is open to free trade is good for the whole world. In Europe, the chances that the General Agreement on Tariffs and Trade will succeed increased when the United States said "yes" to NAFTA. And it made possible Clinton's symbolic success at this week's Pacific economic summit in Seattle with leaders from Asia's economically strongest nations, including Japan, Indonesia, Taiwan and South Korea.

The Pacific summit has not led to any concrete proposals; it was not meant to. What it has achieved is a general sense of potential cooperation that was not there before — again, no small matter. Clinton's emergence as an international leader bodes well for an improvement of sagging economies around the world. Certainly, his stature in Seattle is tied to his success on NAFTA, which showed solid influence over Congress. □

The world's money

Industrialized countries still in recession would be helped by their own version of NAFTA for Europe.

How will we all pay our bills? That, on behalf of British taxpayers, is the question the British Chancellor of the Exchequer, Mr Kenneth Clarke, will attempt to answer next Tuesday (30 November). He has a particular problem — a budget deficit of £50 billion a year — but he is not the only one in that predicament. Most governments in other industrial countries have gone as far as they dare in putting off paying their bills. The big spenders, notably the United States and Germany, are in similar case. It will be interesting to see how far the US Congress's recent carefulness with cash has been offset by the promises to spend more that President Bill Clinton has had to make to get the North American Free Trade Agreement (NAFTA) legislation through the House of Representatives. Germany meanwhile is having to tax and borrow to pay for the rehabilitation of the eastern *Länder*. Where will it all end?

In the short run, in Britain, there will be higher taxes. Some tax increases, equivalent to 6 per cent of the annual deficit, are already written into law. More are needed if the long-term risks of a huge accumulating deficit are to be



avoided. But that is a nasty choice to make. Thanks to a 15 per cent devaluation last September and, since then, interest rates lower than elsewhere in Europe, the British economy has been expanding, but at a rate of 1 per cent a year only. Is there not a danger that higher taxes will snuff out this modest 'recovery'? Of course, money transferred direct to the government from people's pockets cannot be spent on goods and services whose consumption would help to generate or to sustain economic activity. The British government, as any other would, will instinctively shrink from the unpopularity that would bring. Putting off the evil day is a familiar stratagem.

On this occasion, not just in Britain but also in the rest of Europe as well as the United States, the present economic problem is qualitatively new, not just a simulacrum of previous recoveries from recession. There are ample signs of that. Unemployment remains high in the United States and Britain, signs of economic recovery notwithstanding. Elsewhere in Europe, the dole queues will continue to grow for some time yet. The economic community will begin its life as the European Union with more than 20 million people unemployed. People are being turned out of schools and colleges with no work to go to. They are alarmed that there may never be any for them. Why should it be like that?

An important part of the explanation, surprising though it may seem, is the ending of the Cold War. Why should such a welcome development cause such widespread trouble? First, the reductions of military budgets that have been the chief elements of the 'peace dividend' so far have thrown people out of work; California thinks of itself as particularly hard-hit. Second, there has been a substantial diversion of resources from the industrialized West to the newly independent countries of Central Europe and further east. In Germany, where much of the extra cost shows up directly as higher taxes, interest rates higher than they would otherwise be are a measure of the scale on which funds have been borrowed to finance the eastern *Länder*. Official and commercial transfers of investment funds from the West to points still further east probably exceed \$100 billion a year — and are even so inadequate. But both effects are deflationary: the cost of making explosives is a factor in every country's gross national product, investment in the East is possible only at the expense of investment or consumption (or inflation) in the West. The surprise is only that these connections had not been foreseen by the money managers.

Simple oversight does not account for that neglect. There are alternatives, but they are mostly unpalatable. Why not, for example, think of helping the East stand on its own feet by lifting tariffs, quotas and simple prohibitions against goods from Eastern Europe and beyond? Then, surely, consumers in the West would benefit from lower prices, while the East would be less dependent on handouts? There are many reasons, of which the chief is that the pattern of trade with which the industrialized West became familiar during the Cold War would be disrupted. But, in the long run, that is going to happen anyway. Why not anticipate the process?

The answer is: because of all the insubstantial reasons

advanced in the past few weeks by President Bill Clinton's opponents on NAFTA (see above). Especially in agricultural produce and steel, but also refined aluminium and mechanically engineered products, the East would have a temporary advantage. One result, of course, would be job losses in farming and steel-making, but the other side of that coin would be a substantial boom in the provision of the goods and services desperately in short supply in the East. And it would not be a zero-sum game. The combined economic activity of the East and West would be enlarged, to mutual benefit.

The European Union has a particular problem: the Common Agricultural Policy, which costs Europeans \$30 billion a year in taxes and roughly the same in increased food prices. That is a huge deflationary influence in its own right. It is also why the question of the free import of farm products from the East is not discussed. It is true that the European Community has reached an understanding with the three principal Central European states on farm imports, but always within the context of the same agricultural policy. Special help would also be needed to see the groups chiefly affected through a period of adjustment. But that would be far better, and cheaper, than sticking with a policy that seems designed to maintain the pauperization of Western and Eastern Europe for as long as possible. □

Examination farce

There is calculated perversity in the British government's latest policy on school examinations.

FOR many years it has been plain that the curriculum of British secondary schools does not provide a general education. Part of the trouble is that the requirements for entry into universities include good grades in a school-leaving examination for 17–18-year-olds called A-level. Academically inclined students usually follow two-year courses at school that are rehearsals for their later studies, and which are specified in some detail by university departments (especially in science).

This system, with the remarkable precocity it demands, has been called the 'jewel in the crown' of British school education. It is also one of the chief reasons for the flight of young people from science. But, in the past few years, after long negotiation, a modest broadening of the curriculum has been agreed; students may 'take' five rather than the usual three 'subjects' and still win a place at university. (It remains to be seen how many students or even universities will take that option seriously.)

But then what happens? Mr John Patten, in charge of the Ministry of Education, announced (two weeks ago) that there will be an extra grade at A-level, labelled A*, intended to distinguish the better students from all the others. The predictable result will be to undermine the broader option for school-leavers reinforcing inappropriate specialism. The jewel will no doubt shine even more brightly in the crown, but at the expense of young people and their education. □

Britain puts industrialists into the driving seats of science policy

London. The British government has taken an important step towards its goal of putting the 'user community' in charge of the future direction of British science by appointing top industrial figures as the part-time non-executive chairmen of the two new research councils being set up following the publication of its white paper (policy document) on science and technology in May.

It has also appointed Sir John Cadogan, the former director of research at Britain's largest petrochemicals company, British Petroleum (BP), to the new post of director-general of research councils. This is a full-time position within the Office of Science and Technology reporting directly to William Waldegrave, the cabinet minister responsible for science, who announced the appointments last week.

Alan Rudge, managing director of development and procurement at British Telecom (BT), is to be the chairman of the Engineering and Physical Sciences Research Council, the body which is to take over the more technology-orientated research fields of the Science and Engineering Research Council (SERC) when the latter is abolished in its current form next year.

At the same time, Peter Williams, chairman of Oxford Instruments Group plc, will become the chairman of the Particle Physics and Astronomy Research Council (PPARC). This is the body scheduled to take on the SERC's responsibilities in these two fields.

Still to be announced are the full-time chief executives of the two councils. At present, the functions of chairman and chief executive officer are combined at the SERC in a single position; the new arrangements have been introduced as part of the white paper's strategy for linking the science base more closely to the goal of wealth creation.

Wary of concern among scientists about industrialists being given responsibility for a significant part of the science base, both Rudge and Williams emphasize that they are committed to the value of long-term research. Each points out that his company relies heavily on close links with university researchers.

But both are also keen that the value of such research should be increasingly viewed from the point of view of the potential consumer — an approach which Rudge, for example, used as the central theme in his restructuring of BT's research strategy in the late 1980s.

Similar views are also expressed by Sir

John Cadogan, the new director-general of research councils. Currently visiting professor of chemistry at Imperial College, London, he spent 13 years directing the research activities of BP, where he was responsible for the biggest industry-university collaborative research programme in the UK.

Cadogan was well known inside BP for his efforts to demonstrate the value of research to senior executives through the wide-

spread use of cost-benefit analysis by its operating divisions. In particular he persuaded them to identify the precise contribution that company scientists had made to their divisional profits.

He claims that this approach succeeded in convincing management of the importance of the scientists' work. At the same time, it made the scientists aware that their activities were likely to be judged by their value as assessed by the operating divisions. "The failure of previous attempts at such cost benefit analysis was that it was always done by the researchers," he says.

Cadogan says he is aware that this approach may not be appropriate for all areas of government-funded research. But it does have a value, he says "when you have a bottom line". And he adds: "There is no reason why you should not look at it."

Rudge is an electrical engineer by training who, after a spell teaching at the University of Birmingham and in the United States, joined the telecommunications industry in 1979 as managing director of ERA Technology Ltd, joining BT as its director of research and technology in 1987.

At BT, already in the full throes of privatization, Rudge was associated with a wide-ranging restructuring of its research activities. This involved in particular bringing its laboratory scientists — including those supported on contracts in universities — directly in contact with product managers, often for the first time.

Rudge says that this experience taught him the importance of encouraging links between the producers and the users of

research at an early stage in its development, and of making scientists aware that, without wealth creation, there would be no funds for their work.

Williams, a solid-state physicist, also spent a period in academic life — as a lecturer at Imperial College, London — before joining Oxford Instruments, one of Britain's most successful instrument manufacturers specializing in particular in superconducting technologies, in 1982. He became chief executive in 1985, and chairman in 1992.

He, too, argues that wealth creation is a prerequisite for a healthy science base. But whereas BT, in its submission to the government during consultation on the white paper, explicitly argued that "resources should continue to be directed away from 'big science' towards technologies

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS



New team at the top: Sir John Cadogan (formerly of BP), Alan Rudge of BT and Peter Williams of Oxford Instruments.

Hubble stakes are high for NASA's future

Washington. At around 0500 Eastern Standard Time next Wednesday, 1 December — last-minute delays notwithstanding — the US National Aeronautics and Space Administration (NASA) will launch a mission with a public profile unmatched since the Apollo programme of the 1960s. Its success or failure could have almost as much bearing on the troubled space agency's long-term prospects.

The countdown has already begun on the seven-astronaut shuttle mission to repair the Hubble space telescope, and NASA is putting a brave face on its impending trial. "We are not afraid", the agency's chief, Dan Goldin, told a Senate hearing last week. "If we fail,

we will dust ourselves down and do it again."

But the stakes are uncomfortably high for the agency, particularly because the mission's five-day schedule of complex repairs by astronauts working outside the spacecraft makes it one of the most complex ever attempted.

"This is an exceedingly difficult mission, and NASA is under a cloud right now, so it is going to be closely watched by a lot of people," says Frank Gilbert of the National Space Society in Washington. If the mission fails, he says, the scientific and aerospace communities may be understanding; but he warns that the media — and NASA's enemies in Congress — will be

ready to pounce.

The Hubble repair mission takes place as NASA is fighting a losing battle to fend off further budget cuts. Last week Goldin, who since his appointment last year has taken on the mantle of an outsider determined to cut down the agency's bureaucracy, pleaded with senators to desist from further cuts.

"We have cut and cut and cut the NASA programme", he told the Senate science, technology and space subcommittee. "Now we have reached the point where the prospect of additional cuts is of grave concern."

Goldin said that NASA's planned five-year budget had fallen from \$96 billion to \$78 billion in 18 months, and could now fall to \$71 billion. The budget was being changed "every two or three months", he said, calling on Congress to give him a stable amount of money to plan with.

Ken Ledbetter, programme manager for Hubble at the Goddard Space Flight Center in Greenbelt, Maryland, declines to comment on whether he is under special pressure from his bosses because of the importance the mission has assumed.

Ledbetter says that the agency has planned for as many contingencies as it can think of.

Practising optical surgery in space

Paris. Astronomers will be holding their breath on 3 December when a robot arm snatches the bus-sized Hubble Space Telescope (HST) into the cargo bay of the shuttle Endeavour. Astronauts will then set about correcting Hubble's blurred vision, replacing scientific instruments and servicing the telescope's on-board systems to keep it running until the shuttle's next visit in 1997.

Six-hundred kilometres above the Earth, Hubble has a clear view of the Universe in both the visible and ultraviolet range. Its 2.4-metre primary mirror is not big by modern standards. But this vantage point means that Hubble should outperform the largest land-based telescopes, giving it ten times better resolution (the ability to distinguish the two rear lights of a car at 10,000 km) and thirty times greater sensitivity (able to detect a star that is a billion times fainter than the faintest visible to the naked eye).

But the first images from the telescope's Wide Field/Planetary Camera (WF/PC) and Faint Object Camera (FOC), three years ago, exposed spherical aberration in the telescope's primary mirror caused by a manufacturing error. The consequent loss of sensitivity also affected HST's three other scientific instruments — the Faint Object Spectrograph (FOS), the Goddard High Resolution Spectrograph (GHRS) and the High Speed Photometer.

Moreover, although all Hubble's major systems worked at first, three of its six gyroscopes (which help to point the telescope) and a computer memory board have since packed up. Power has also failed in part of the GHRS.

The solar arrays have also caused unexpected vibrations, which occur when the panels heat or cool suddenly between day and night orbits (once every 90 minutes). Changes to the HST pointing and

control software have partly solved the problem, but the unpredictable onset of vibrations often leads to loss of fine lock during observations.

Some of Hubble's problems were unexpected. But the original mission plan anticipated regular shuttle missions to service parts of the telescope. To this end, all the scientific instruments and critical spacecraft systems have been designed as modules that can be replaced by astronauts.

This first service mission — which is also intended to test the feasibility of in-orbit servicing (see above) — is planned to last 11 days. Five six-hour spacewalks are planned, and contingency plans allow for a further three.

"Minimum success" would involve fitting three new gyroscopes, and either the replacement of the Wide Field/Planetary Camera or fitting the COSTAR optics (Corrective Optics Space Telescope Axial Replacement). "Maximum success" would involve fitting the WF/PC, COSTAR, six new gyroscopes, the replacement solar arrays (thermally shielded, and redesigned to buffer thermal expansion and contraction) and one magnetometer.

The COSTAR re-imaging mirrors will compensate the FOC, FOS and GHRS for the aberration in the primary mirror (the space occupied by COSTAR will require sacrificing the High Speed Photometer). The replacement WF/PC is equipped with its own system to correct the spherical aberration.

The first aberration-corrected pictures will not be available until around two months after Endeavour touches down at the Kennedy Space Center in Florida on 12 December. If all goes well, astronomers' attention will then turn to using the Hubble Space Telescope's long-awaited potential.

Declan Butler

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Good vibrations: new solar arrays should reduce problems encountered previously.

"We've constructed the repair tasks to be modular, we've cross-trained the crew, we have done so many things," he says. "But space is the frontier, and when you push the edge of the frontier, you sometimes hit the unexpected."

Indeed, it is the tendency of spacewalks to encounter unexpected problems that makes many observers sceptical about whether NASA can successfully complete the necessary 30 hours of repair work in the allotted time. NASA officials reply that the 11-day mission could be extended by as many as three days, and that a sixth day of

repair work may be added.

The officials also stress that the success or otherwise of the repair will not be immediately apparent. Alignment and calibration of the equipment installed to improve Hubble's resolution will not be complete for seven weeks, and a full programme of science using the telescope is not expected to resume for 14 weeks.

If an additional shuttle mission is needed to complete the repair, it could take place in between five and thirteen months time, depending on the tasks needed, says Joe Rothenberg, flight projects manager. Officials on Capitol Hill say that Congress would probably approve such a mission in order to rescue the \$3 billion already invested in Hubble.

But the ramifications of technical problems on next week's mission would be much broader, influencing in particular Congress's perception of NASA's ability to construct the space station, on budget, through a series of shuttle missions. "If this goes wrong it will be very serious", says Washington-based space consultant Wilfred Mellors.

Colin Macilwain

Multimedia patent sparks protest

Washington. A small California company which has won a broad patent covering virtually all types of multimedia computer software is expecting a torrent of litigation from rival companies before it attempts to extract licence fees from them next June.

These companies, as well as several patent lawyers, have attacked the US Patent and Trademark Office for issuing such a broad patent, and called for its procedures to be changed to allow outsiders to comment on patents before they are awarded.

The US patent has been awarded to Compton's NewMedia of Carlsbad, California, and raised both eyebrows and tempers when its existence was revealed at a computer exhibition in San Diego, California, last week. The patent appears to cover almost any type of database capable of storing video and audio information along with computer data.

"Any patent invites some sort of challenge", says Pat Meier, a spokeswoman for the company. "Initially people were infuriated by the announcement, but now they are settling down and asking what it really means."

Compton's NewMedia will seek a three per cent licence fee from rivals after next June, Meier says, although those who make a distribution agreement by then will pay less. The privately owned company has 150 staff, and has pioneered the development and retailing of multimedia, computer-based encyclopaedias over the past ten years. But revenue from the patent may dwarf income from its current activities.

Colin Macilwain

Plant biologists sow seeds of EU funding experiment

Paris. The European Commission may delegate management of much of the European Union (EU)'s proposed ECU13.1 billion (US\$14.8 billion) fourth Framework programme to the scientific community. The decision will depend partly on the outcome of a test case within the EU's third Framework programme, where a consortium led by the John Innes Institute in Norwich (UK) and the Max Planck Institute in Cologne (Germany) will manage an ECU24 million plant molecular biology project.

Technically speaking, the Commission will contract the project to a European Economic Interest Grouping (EEIG) comprising the John Innes and Max Planck institutes only. But the institutes say this is only to simplify legal arrangements, and that the EEIG will delegate power to a board made up of leading scientists from several European countries.

This board, known as AMICA (A Molecular Initiative in Community Agriculture), was formed four years ago. It aimed to persuade the then research commissioner, Filippo Pandolfi, that it should organize, seek proposals for and peer review all Commission research in plant molecular biology.

Last year, the Commission decided that this could not be done, given its legal responsibility to European taxpayers to decide how money is allocated. But the Commission invited AMICA to submit, as a trial, its own proposal for the EU's Priority Technological Programme (PTP) on Plant Molecular Genetics for an Environmentally Compatible Agriculture.

AMICA subsequently sent out more than 500 applications for peer review, and selected 116 projects. But the Commission's own subsequent peer-review process eliminated half of these. The Commission also funded about 50 applications out of 500 received separately from small independent research consortia.

AMICA's original proposal was dead, but it was subsequently asked to manage the entire PTP. The Commission wanted to test devolved management, not least because a recruitment freeze meant that increases in its scientific staff would not keep pace with the proposed doubling of the EU's research budget in the fourth Framework programme.

After long negotiations, AMICA-EEIG

persuaded the 50 or so independent groups that the Commission had chosen to come under its umbrella. As a result, AMICA-EEIG will now coordinate and evaluate all 117 laboratories taking part in the PTP; previously, EU programmes consisted of many small consortia, each linking on average five laboratories, and were managed by commission officials.

Jeff Schell, director of the Max Planck Institute, sees AMICA-EEIG as a model for EU involvement in science. He said that coordinating efforts in national institutions meets the principle of subsidiarity. Richard Flavell, director of the John Innes Institute, says a major aim of the programmes is to use the size and diversity of the participating groups to add value to national and international programme on plant science.

One Commission official says if AMICA-EEIG succeeded, similar structures were likely to be set up in other research fields within the fourth Framework programme.

However, some researchers fear that devolved management, which they welcome, could be the first step towards devolution of the power to determine which projects receive EU funding and to shape future programmes.

Jean Dénarié, from the Centre National de la Recherche Scientifique Laboratory in Toulouse (France) criticizes what he describes as the "concentration of power" in AMICA-EEIG, and in particular its initial attempt to select proposals.

Dénarié also argues that putting a few leading scientists in control of all EU-funded work in a particular field will lead to poorer countries being given less. He adds that the European Commission is best placed to provide independent control over who gets funded what. "Community research is not only about funding Europe's best laboratories, but about developing the scientific potential of those countries which lag behind," he says.

Schell agrees that a science-driven system would provide a "selective advantage to those institutions and countries that do a good job in science". But he points out that the Commission had earmarked ECU1.5 million of AMICA-EEIG's programme for training programmes in Greece, Portugal and Ireland. "The jury is still out" on whether poorer countries would accept AMICA-EEIG, he says.

One Commission official says the question of delegating control of research programmes to groups of scientists is "not open to discussion". But he admits that it could happen "if there were sufficient safeguards that the ultimate responsibility lay with the Commission".

Declan Butler



Richard Flavell:
"adding value".

Japan builds ivory towers among its windswept hills

Kansai Science City. Japan's second major science city, after Tsukuba science city north of Tokyo, is beginning to take shape in the Keihanna hills between Kyoto, Nara and Osaka, about 200 miles west of the nation's capital.

Until recently Kansai science city, in which local and central governments and private industry plan to invest ¥4–5 billion (US\$37–47 million) over the next few decades, was nothing more than a university, a couple of colleges and a handful of institutes widely scattered in the Keihanna hills.

This year, however, has seen the opening of a central service complex and four new institutes — three in the past few weeks alone — in a small district intended to be the heart of the new city.

On 10 November, the central research laboratory of the Research Institute of Innovative Technology for the Earth (RITE) began operations in a futuristic building powered in part by solar cells. The organization is being funded generously by both private companies and the Ministry of International Trade and Industry to develop environment-friendly technology.

The RITE laboratory has 35 researchers, and that number will double over the next five to 10 years. The central building has a slanting roof covered in solar cells. On either side extend two long single-storey wings with large bay windows looking out onto a moat-like pond and a huge landscaped garden.

Next door is the International Institute of Advanced Studies (IIAS), the 'brain centre' of the city, which opened last month and is modelled on the Princeton Institute of Advanced Studies in the United States (see *Nature* 365, 482; 1993).

Five hundred metres away is the towering Advanced Telecommunications Research Institute International (ATR), which was the first institute to open in the city in 1989. It is funded with the dividends of government-held shares in Nippon Telegraph and Telephone Corporation.

Facing ATR is the Keihanna Plaza, opened last March, which acts as a central service complex. The world's largest sundial graces a square in front of the complex. And last week, Matsushita Electrical Industrial Company opened its research institute in the same district. It will have a total staff of about 300 when it reaches full operation.

When complete, the city will be composed of 12 such districts or 'clusters', separated by wooded hills. Each district is quite small. The Seika-Nishi Kizu district accommodating the Keihanna Plaza and the new institutes is one of the largest, yet still covers only 500 hectares.

The 12 clusters will be linked by a computer network centred on the Keihanna Plaza. And planners hope eventually to link all the clusters with a grid system of roads.

At present, however, it is quite difficult to travel between and even within districts. This aspect of Kansai is reminiscent of Tsukuba which, after 20 years of development, still lacks a railway station.

The Keihanna Plaza, which has a concert hall, a hotel, a general store, a travel agency, a bank, a post office and a noodle restaurant that is transformed into a bar at night, is an attempt to make researchers feel more at home.

But there is still the impression of being in the middle of windswept hills far from civilization. And many of the city's current 6,000 researchers have chosen to live in Kyoto, Nara or Osaka, 10–30 km away.

There are new housing estates adjacent to the research districts. But house prices were driven to astronomical levels of more than ¥100 million yen (\$1 million) during the 'bubble economy' at the end of the 1980s. Prices have since fallen. But they are still beyond the pocket of most researchers.

It is therefore hard to see how the city will be developed into the "organic whole" that its developers have promised. Rather, there is a danger that the lavishly funded institutes will become ivory towers in pleasant surroundings.

David Swinbanks

Russian foundation challenges power of the Academy

Moscow. Vladimir Fortov, the new director of the Russian Foundation of Basic Research (RFBR) who took over the helm in June, is trying to persuade the Russian government to double the foundation's budget.

Fortov says he is confident that the foundation, set up in April 1992 along the lines of a Western science funding agency, can use funds far more effectively than the Russian Academy of Science. He says that this is primarily because much of the money channelled to the academy is spent on administration; the RFBR limits overheads to 20 per cent of its budget.

Under Fortov's leadership, new regulations were approved last month which considerably reduce the influence of the foundation's council on decisions by its expert panels on the allocation of research funds. Previously the council had been criticized for allocating a high proportion of grants to its own praesidium members.

The structure of the RFBR is also to change. In particular, an executive committee is being set up to carry out some of the functions of the council between its meetings. And a growth in the number of advisers means that each grant proposal is now reviewed by two experts rather than one.

The foundation's work has been developing rapidly since Fortov replaced Andrei Gonchar, the vice president of the Russian Academy of Science, last summer. This followed a power struggle in which Gonchar was told that he could not continue to hold both positions (see *Nature* 363, 662; 1993).

Despite the general lack of money in Russia, the foundation has been successful in obtaining from the government most of the money it had been promised. This year it expects to receive about 20 billion rubles (US\$18 million).

Nevertheless some problems still exist. For example, most of the foundation's research projects have been planned to last for three years. This means that about 80 per cent of the programmes now being funded will need funding for the next year — and that next year's competition for funds could end up providing support for only 600 out of an expected 15,000 applications.

The foundation has therefore increased the portion of its budget allocated to research grants from 60 to 75 per cent. This should be sufficient to guarantee funding of 2,200 new projects.

If the budget of the RFBR were to be doubled, its financial power would equal that of the academy, which receives between 10 and 11 per cent of the state annual budget for science. The academy is therefore likely to oppose any such action.

Vladimir Pokrovsky

China urged to treat young scientists better

Peking. A nationwide survey of China's leading young scientists has found that the majority are concerned about low pay, difficult working relationships and poor living conditions.

The survey was carried out by the China Association of Science and Technology, which interviewed 291 winners of the "China Young Scientist Award" working in research laboratories and universities.

The numbers were evenly balanced on

the question of whether the scientists felt they were attaining their full potential. Over two-thirds of the scientists agreed that their institutions cared about their work. But three quarters said that they had bad experiences of interpersonal relationships, quoting jealousy, in-fighting and professional rivalry.

As a result of the survey, the association has suggested that research institutions introduce a "more benign attitude" towards young scientists.

You Qin Li

Court backs critic of Italian jobs 'carve-up'

Munich. In a decision widely seen as a blow to the traditional power of 'old-boy networks' in Italian universities, a Genoan researcher has successfully challenged the results of a nationwide recruitment examination held to recommend the appointment of individuals to full professor.

Gian Franco Gaetani of the University of Genoa Medical School was rejected in 1988 by one of the 300 committees set up in that year to make new university appointments under Italy's nationwide centralized competition, known as the *concorso*.

Gaetani complained to an administrative court, claiming that the proper procedures for assessing candidates were not followed, and that the committee that refused him promotion failed to give adequate weight to his scientific achievements. Five years later, his complaints have been upheld, and each of the five appointments made by the committee have been declared null and void.

The decision has been welcomed by many Italian scientists, who have been campaigning to end the country's unpopular system of academic promotion. They claim that this system is often manipulated by the country's academic elite, accusing it of 'cronyism' and of using the appointments process to exchange favours.

Openings for professorships, assistant professorships and associate professorships are not advertised by universities in Italy, but are offered through the *concorso*, which

contributions made by candidates to collaborative papers whose authors also included a member of the committee.

The court did not try to assess the scientific achievements of individuals themselves. But it did decide that the committee had been unfairly selective in its use of information from applicants. (Gaetani had claimed that his membership of important international scientific committees was ignored.)

This is the first time all the decisions of a particular committee have been declared invalid. Forty individuals have bought cases against the 1988 *concorso*, but so far only Gaetani has won; seven cases have been dismissed, and the remainder are pending.

The five professors promoted on the recommendations of the haematology committee now return to their former rankings, while the committee decides whether to appeal to a higher court. If the administra-

tive's court's verdict stands, the posts will be reopened through national competition.

As the committee has a year to decide whether to appeal, the promotions will probably miss the round of the *concorso* now being evaluated.

The haematology committee has another problem. When it was set up for the current *concorso*, it included some of the first recent appointees whose status as full professor has been removed. Rather than risk a second challenge to the validity of its decisions, the whole committee has been dissolved, and will be reappointed later in the year.

Changes in the whole promotions systems to Italian universities are under discussion. But unless they are introduced in the near future, the five appointments will have to wait until the next *concorso*, which will not take place for several years.

Alison Abbott

WHO widens focus of AIDS research

Paris. The emphasis of AIDS research on therapeutic antiviral drugs designed to attack the late stage of the human immunodeficiency virus (HIV) in infected persons produces exciting science; but it may not be the most effective way to avert an AIDS epidemic.

This was the underlying message of the decision last week by the World Health Organization (WHO) to support research into foams and creams designed to inactivate HIV in a woman's vagina.

WHO said that vaginal microbicides could revolutionize efforts to prevent heterosexual transmission of HIV by protecting women, especially in developing countries, who cannot ensure that their sexual partners use a condom.

Michael Merson, executive director of WHO's Global Programme on AIDS, said the condom remained the main defence. But with two women infected with HIV every minute, "we need a new method for women to protect themselves".

Rudi Pauwels from the Rega Institute for Medical Research in Belgium welcomed the organization's commitment to an area that he said had been neglected by mainstream antiviral AIDS research. "We have been focusing on the infected person, whereas the epidemiological data tells us we should be looking at how to prevent it," Pauwels said. "I'm not saying that we should drop vaccine research and related activities, but we need to look at all the options."

Prospects for progress towards a vaginal microbicide are uncertain. There is some evidence that existing spermicides, such as the macrogel ether nonoxonyl 9 (N-9), which has been used for more than 40 years,

inhibit HIV *in vitro*. But Pauwels said that not enough is known about the effect of viricides have on genital transmission of HIV.

Joep Lange, head of clinical research with the WHO's Global Programme on AIDS, admitted that the concept may be flawed. "But it's worth a try," he said.

Certainly, the prerequisites for a safe topical preparation may be less stringent than for a systemic drug. Potential candidates include polyanions, such as dextran sulphate, and other inhibitors of virus adsorption/fusion. The WHO also believes that promising antivirals abandoned because of problems of bioavailability or toxicity may be suitable for vaginal preparations.

Lange said the WHO's commitment would encourage research into the use of topical preparations. The WHO would also stimulate the interest of companies by carrying out and paying for clinical trials.

According to Lange, the WHO label would lend credibility to such trials. He said that many developing countries are increasingly worried about being exploited by pharmaceutical companies trying to dump products, or entrepreneurs marketing dubious wonder-cures.

Lange also emphasized that all safety tests on vaginal microbicides would be carried out first in low-risk populations in industrialized countries. Some companies have tested existing spermicides for HIV-inhibition in high-risk populations in Kenya and Senegal, only to discover that they caused vaginal lesions. The treatment may thus have inadvertently — but unethically — exposed subjects to increased risk of HIV infection.

Declan Butler



takes place every four or five years. More than 10,000 applicants were considered in the 1988 competition; by law, each should have been assessed on the basis of written applications alone listing his or her scientific achievements, including publications and service on international committees.

In November 1989, the haematology committee recommended the appointment of five candidates. But Gaetani subsequently showed that their track records, based on indices of the impact of scientific publications listed in the *Science Citation Index*, were lower than those of four of the ten losers (see *Nature* 353, 10; 1991).

The court has now declared that the five appointments are invalid, arguing that the committee failed to distinguish the precise

Germany's 'blue-list' gets tougher judgement

Munich. Eighty-two research institutes in Germany, known collectively as 'blue-list' institutes, are to be given a stronger umbrella organization and a formal system of evaluation that will make it more difficult for poor performers to use political backing to protect themselves against closure.

Blue-list institutes — named for the colour of the paper on which the original list was printed — carry out a range of research activities, from low-temperature plasma physics to tropical medicine.

The cost of each institute is shared by its home *Land* and one of ten federal ministries. But they have never fitted comfortably within the German basic research system, which has three major pillars — the universities, the Max Planck Gesellschaft (MPG) and the national research centres.

Difficulties started after the reunification of Germany, when 34 establishments of the former East Germany were accepted as blue-list institutes on the recommendation of the science advisory council, the Wissenschaftsrat. Each was felt to be too small to become a national research centre; but neither the MPG nor the universities — as had been hoped — were prepared to absorb them.

As a result, the number of such institutes increased from 48 to 82 and their budget, currently DM1.16 billion (US\$700 million) a year, began to rival that of the MPG (see *Nature* **363**, 482; 1993). Critics say there are now too many blue-list institutes, and that they should be made to fit into the conventional research establishment. But Gerhard Neuweiler, head of the Wissenschaftsrat, has long disagreed.

He has wanted their small umbrella or-

ganization, the AG-BL, which was set up in January 1992, to assume similar powers to those of the MPG, which controls not only its own budget and the scientific direction of the society as a whole, but also, through a rigorous system of evaluation, the standard of research in its individual institutes. This, he says, would allow money to be shifted away from institutes that perform badly.

Until now, evaluations of blue-list institutes have been carried out on request by the Wissenschaftsrat. Neuweiler says that in the past, negative evaluations have been met with political pressure either from the host *Land*, keen to retain research institutes that enjoy federal support, or from the relevant ministry, reluctant to lose the prestige of its own research institute.

But there has also been resistance to the idea that the blue-list institutes should have more power and status. The Wissenschaftsrat had to accept a compromise. But this will still allow the AG-BL to enlarge its secretariat and set up a permanent base.

The AG-BL will probably be based in Cologne, the home of the Wissenschaftsrat.

It is now planning to set up a scientific section to guide its research portfolios, and to carry out systematic evaluations of its institutes through the Wissenschaftsrat, starting in the new year.

In order to reduce political influence on decisions about which institutes should be supported, the Wissenschaftsrat has introduced a new procedure under which the original scientific assessment of each evaluation committee will accompany the final recommendation of the Wissenschaftsrat, and will therefore be made public. The Wissenschaftsrat is setting up a standing committee to organize evaluations, a move which will double its workload.

Neuweiler hopes that, with the new arrangements, arguments over the continued existence of the blue-list institutes will cease. "Two years ago the Wissenschaftsrat promised a good future to scientists in blue-list institutes in the new *Länder*," he says. Neuweiler adds: "We can't say now that we have second thoughts". But he admits that there remains some "hidden pressure" to say just that.

Alison Abbott

Trial deaths prompt tighter rules

Washington. The Food and Drug Administration (FDA) is to tighten up its clinical drug trial regulations following a trial of a treatment for hepatitis-B carried out by the National Institutes of Health (NIH) earlier this year in which five patients died of liver failure (*Nature* **364**, 275; 1993).

New regulations being drawn up by the FDA will make drug companies who sponsor trials responsible for gathering informa-

tion on side effects from all previous safety and efficacy trials of the drug in question, and for looking for patterns of toxicity.

The regulations were recommended in a report by an FDA task force into the immediate lessons of the trial, which used the drug Fialuridine. An internal NIH investigation has already ruled that NIH procedures were fully complied with. But in response to Congressional concern, health secretary Donna Shalala has asked the Institute of Medicine to conduct a full, independent review of the trial.

The five died when the drug was used on fifteen patients in a trial conducted by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and sponsored by the company Eli Lilly.

The report examined data from six previous clinical trials of Fialuridine, and found that problems arising in them were attributed to causes other than the drug's toxicity.

But the report finds that a "worst case" analysis of all deaths and adverse effects associated with previous trials could have detected the drug's toxicity before the fateful trial took place. The FDA report also notes that the natural optimism of investigators and sponsors can lead them to miss evidence of toxicity.

The task force, having recommended changes to existing regulations, is now investigating whether the NIDDK research team, led by Dr Jay H. Hoofnagle, breached old FDA rules in its conduct of the trial.

Colin Macilwain

Biologist to head US Forest Service

Washington. US conservationists have enthusiastically welcomed the appointment of a wildlife biologist, Jack Ward Thomas, as chief of the Department of Agriculture's Forest Service. But representatives of the timber industry are furious that Thomas, who has worked with the Forest Service as a wildlife biologist for 27 years, has been chosen to fill a key post previously held by career forest managers.

As a prominent researcher associated with moves to restrict tree felling, Thomas has been an active participant in a passionate debate over forestry policy, especially in the northwestern United States. Even before his promotion, he had travelled in the company of bodyguards and had received a number of death threats.

In 1990, Thomas chaired the panel that revealed the threat posed by the activities of the timber industry to the survival of the spotted owl in the northwest. His appointment seems certain to mark a clear shift in the Forest Service's first priority, from timber production to nature conservation. "Thomas represents a new generation of managers, fully committed to maintaining biological diversity", says Stuart Pimm, professor of ecology at the University of Tennessee.

Jack Ward Thomas

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Anneli Høglund

South African agency faces new priorities

Cape Town. F. W. de Klerk, president of South Africa, and Nelson Mandela, president of the African National Congress (ANC), last week agreed terms for a draft constitution, clearing the way for South Africa's first open elections. For institutions such as the Council for Scientific and Industrial Research (CSIR), the main channel of government funds for applied research, next year's election results are awaited with some trepidation.

The CSIR performed well in a recent ANC review of science and technology institutions. But it is undergoing a period of evolution and uncertainty that began with a restructuring in 1988. The council is changing its focus from pure and applied research to research orientated specifically towards technology and its development.

In the 1980s, South Africa had a lucrative market in defence and other strategic industries. Government spending on research and development has, however, since dropped from 0.82 per cent to only 0.55 per cent of the federal budget. After next year's election, the government's spending priorities will almost certainly be different.

Brian Clark, the president of CSIR, believes the council's focus will remain on technology, and sees this being directed towards three goals: making South African industry more competitive in international markets; promoting local development; and assisting in government decision-making, for example through advising on the environmental impact of new developments.

Between 1988 and 1993, the new CSIR had its parliamentary grant increased by only 7 per cent, from R192 million (US\$45 million) to R205 million, well behind inflation. As a result, government funding has now dropped from 55.5 to 46 per cent of the council's income. The balance is made up of income derived from research contracts and investments and royalties from patents. The former has risen from 35 to 46.5 per cent — exceeding the parliamentary grant for the first time last year.

The council has had increasing success in attracting foreign funds, as well as support from the private sector. Six per cent of its contract income last year came from abroad, compared with one per cent in 1988.

In contrast, research contracts from both the civil public sector and the defence industry have dropped sharply and are expected to fall further this year. There is a question over whether the CSIR can continue to increase its income from the private and foreign sectors, and whether it can hold onto its parliamentary grant.

This year, the council's grant was increased by 16 per cent — the first increase in real terms since its restructuring. The decision prompted speculation that it was being 'bailed out' by government. But Clark denies this, explaining that most of the new money was part of a special allocation earmarked by the state for additional employer contributions to the CSIR's pension fund.

The council fared better than many other bodies in a review of South Africa's science

and technology system compiled for the ANC (see *Nature* **362**, 384; 1993). However, the review criticized the council's emphasis on meeting the needs of big business, pointing out that only a third of CSIR contract income is derived from small and medium-sized companies.

In response, the CSIR has launched a special initiative in technology transfer to these smaller enterprises. But some scientists still feel that the council is not much more than a giant consultancy, and as such has no real claim to a state subsidy.

Jennifer Thomson and Johan Lutjeharms, two former CSIR employees who now hold chairs at the University of Cape Town, claimed earlier this year in the *South African Journal of Science* that "science as a thriving, innovative, intellectual enterprise within the CSIR is dead".

Certainly, the number of publications produced by CSIR employees has declined from more than 300 in 1988 to fewer than 100 last year.

But in the same issue of the journal, CSIR's vice-president, Daan Toerien, argued that, in the context of its new mission, it was more relevant that the number of patent applications filed by the CSIR had trebled over the same period.

Anastassios Pouris, newly appointed director of science and technology policy at the Foundation for Research Development says there is still a great deal of confusion. "The problem with the present science system is that the taxpayer doesn't know whether he is subsidizing research for the private sector, basic research, strategic research, routine testing or even, in some cases, just enabling councils to compete favourably with private consultancies."

Questions have also been raised about the CSIR's continued close links with the military: the South African Air Force recently rejected the purchase of the 'Ovid' aircraft, developed locally with the CSIR, in favour of the Swiss 'Pilatus'.

But Clark denies that the CSIR is tainted by such links. "At an executive level we have a very good open relationship with the democratic movement", he says, adding that what is required is "transparency and openness, not the secrecy of the past".

The main issue facing the CSIR is whether the promotion of technological research and development (R&D) is best achieved by investing state funds in a research council. But in the short term there may be little alternative. There is very little in-house R&D being undertaken by the private sector in South Africa. Nor is there any comprehensive incentive scheme by which R&D spending can be written off against tax, a procedure that some feel could be a better way of promoting technological development.

Michael Cherry

Mandela backs science park plan

Cape Town. Africa's largest research and manufacturing park is to be built on a site in Cape Town. The first construction phase is due to begin in October next year, six months after the country's first democratic elections.

The science complex, to be known as Capricorn Park, is being developed by the British company Science Research Parks Limited, at an estimated cost of R1 billion (US\$230 million). It is expected eventually to provide jobs for 40,000 people at 32 industrial and research sites.

Nelson Mandela, president of the African National Congress, has already thrown his weight behind the project. Addressing the Confederation of British Industry in London last month, he said the park "should contribute significantly to science and technology development" in South Africa.

The goal of the developers is to provide international high-technology manufacturing companies with an opportunity to invest in South Africa now that economic sanctions against the country have been lifted.

Supporters of the planned park claim that the Western Cape — one of nine regions proposed in the new federal constitution — could become the 'Silicon Valley' of Africa, with its relatively highly skilled workforce and its pleasant climate and surroundings.

The region's three universities and two technical institutions (technikons) will participate in activities at the park through the Cape Science Foundation. The park will place a strong emphasis on training programmes and an academic centre will form the central focus of the site.

South Africa has virtually no domestic tradition of privately financed research and development, and the success of the project will therefore hinge on whether international companies are attracted by investment opportunities under the new government. The region already has a 'technopark' in Stellenbosch, less than 30 miles from Cape Town. But only three of its 41 sites have been developed, and a further seven sold, since it was founded eight years ago. **M. C.**

Evolution and religion

SIR — Vaneechoutte¹ criticizes Josephson² for saying: "...the central theme of religion is the attempt to maximize human goodness" and Baidins³ for holding that religion helps people to "make more constructive decisions. . .".

Vaneechoutte prefers "to look at religion as an emergent characteristic, which can arise only after other levels have come to full development". It is only proper to acknowledge that there are informed but diverging opinions on this issue.

"Religion . . . is the opium of the people" (Marx). Not bad, but this does not explain its evolutionary value. Why should people tend to believe in religions regardless of facts? Most Americans still regard the biblical account of creation as more probable than the scientific one.

But where is the selective advantage in believing something that is not in agreement with observations? In the first place, when people are bound by religion they stick together. Poland's identity through the centuries was preserved by its Catholicism. Religion in Russia was not extinguished by 70 years of official atheism and religious oppression.

But there is more. Bondi⁴ remarks "The past as well as the present can leave no doubt that the variety of religions is a calamitously divisive force in human affairs". But of course. That's the whole point of religion; that is what provides its evolutionary advantage. When religion divides "us" from "them", God then tells "us" to kill "them".

In Deuteronomy 20:16, God gives the Hebrews their marching orders: "But of the cities of these people, which the Lord God doth give thee for an inheritance, thou shalt save alive nothing that breatheth." Soon thereafter we find Joshua 6:21 on the fate of Jericho: "And they utterly destroyed all that was in the city, both man and woman, young and old, and ox, and sheep, and ass, with the edge of the sword."

Pascal was surely right when he said: "Men never do evil so completely and cheerfully as when they do it from religious conviction." It's neither easy nor fun to slaughter large numbers of innocent men, women and children; but God says it has to be done, and thereby, one way or another, our genes replace theirs. "Death to the infidel!" Compare the success of the Arab tribes before and after Mohammed, AD 632.

A religion has to be sufficiently believable to attract a lot of converts: "God is always for the big battalions" (Voltaire). Yet it must be sufficiently unbelievable that there are still plenty of infidels — otherwise whom would we have to conquer?

A successful religion must also make

death attractive so that its soldiers will fight fearlessly. In Islam a man slain in battle can expect to dine that evening with Mohammed in Paradise and to be issued with some hours. Christian soldiers get a pretty good deal too, but painted in more abstract terms.

One need not indulge in teleological or theological considerations to understand that in mankind the evolutionary mechanism, which is differential reproduction, is largely guided and powered by religion.

R. H. Good

Department of Physics,
California State University,
Hayward, California 94542, USA

SIR — I have two objections to the views of Vaneechoutte¹ that human superintelligence has left us vulnerable to "endless fear and longing", that we differ from other animals in this respect, that memes such as religion and astronomy are natural defensive measures that help us to feel more secure and in control — and that we often resort to nonmemetic solutions such as drugs and overeating.

First, the dividing line between humans and other animals is not as sharp as Vaneechoutte implies. Studies of great apes and monkeys have demonstrated their vulnerability to serious depression and chronic emotional stress. In chimpanzees (*Pan troglodytes*), for example, a youngster whose mother has died may display behaviour that closely resembles human depression⁵, and in olive baboons (*Papio cynocephalus*) a male's subordinate status may be accompanied by high basal corticosteroid levels and behaviour that characterize the syndrome of chronic emotional stress in humans⁶. As cognitive skills increase in phylogeny, the potential for complex modulation of affect also appears to increase, with no sharp dividing line. As Stephen Jay Gould has aptly stated in a different context, nature abhors boundaries.

Second, evolution may indeed have made us "naturally unhappy organisms", as Vaneechoutte asserts, but I suggest that this is only part of the story. It seems likely that large brains enable us to have complex positive feelings too, such as a sense of personal fulfilment, the elation of pair-bonding, and love. One important variable here is mental health. If one assumes a hypothetical mental health continuum, it is easy to imagine a preponderance of dysphoria and depression at the one end and a sense of well-being and 'happiness'

at the other. Affective states along the entire continuum are products of complex brains.

Gregory G. Dimijian

University of Texas,
Southwestern Medical Center at Dallas,
5323 Harry Hines Boulevard,
Dallas, Texas 75235-9070, USA

SIR — Bondi⁴ presents the "incontrovertible" argument that because the views of those holding different religious faiths contradict each other, then only one at most can be right. It is interesting to note that this is also the traditional Judaeo-Christian belief. Underpinning that religious belief is the conviction that there is such a thing as absolute truth, and indeed that there is also untruth. The very act of doing science is predicated on the same acceptance that some things are true and others untrue. There is no inconsistency between holding the Christian faith and being a scientist.

That people have done, and continue to do, unbelievably cruel and evil things to other people is, sadly, only too clear. That such things have sometimes been done in the name of religion should be a cause of shame in all of us. But they are more a reflection of human nature than of particular religious views. The Holocaust and the burning of one's perceived religious enemies were undoubtedly evil; the Christian perspective is that so too is the social injustice in our country, the domestic violence in our homes and the petty selfishness in our everyday lives.

Though we agree with Bondi that the universal and global enterprise of science can in a unique way unite people of different race, culture and background, we believe that science can never address the problem of evil which is a universal feature of the human condition. Only religion claims to do that.

Robert S. White

Department of Earth Sciences,
Colin Humphreys
Department of Materials Science and
Metallurgy,
University of Cambridge,
Cambridge, UK

SIR — Bondi⁴ errs in implying that, because religions are mutually exclusive, most religious people are wrong and that every religion should therefore be mistrusted (*Nature* 365, 484; 1993). His error lies in presuming that man is the only agent involved in the choosing of a religion — when religion concerns an omnipotent God having a freedom and volition of his own. His argument assumes what it sets out to prove.

Anthony Garrett

Byron's Lodge,
63 High Street,
Grantchester,
Cambridge CB3 9NF, UK

1. Vaneechoutte, M. *Nature* **365**, 290 (1993).
2. Josephson, B. D. *Nature* **362**, 583 (1993).
3. Baidins, A. *Nature* **346**, 693 (1990).
4. Bondi, H. *Nature* **365**, 484 (1993).
5. Goodall, J. *The Chimpanzees of Gombe: Patterns of Behavior*, 66 (Belknap, Harvard, 1986).
6. Sapolsky, R. & Ray, J. C. *Am. J. Primat.* **18**, 1–13 (1989).

Has physics come to an end?

The idea that physics has almost been worked out as a field for interesting investigations is widely current but mistaken. There is as much left to do as has been done.

New Delhi. What is there left in physics? That was the plaintive question this week from a score of New Delhi University's physics faculty, worrying about the difficulty of teaching students anything of value without modern or even adequate equipment. One of their concerns is that even the brightest students, starved of stimulation, lose heart and drift off to other occupations. Another is that they themselves, confined to theoretical studies by shortages of equipment and deprived of the stimulation of stimulated students, are understandably unsure about their subject. The news of the abandonment of the Superconducting Super Collider, which has travelled fast, has not much helped. It is a fair guess that the same lack of spirit is to be found in better provided places.

But, with all the extenuating circumstances, the question resonates uncomfortably with triumphalist statements from the likes of Stephen Hawking that the "theory of everything" is within reach. What is there left in physics, in all ways the field of enquiry from which reductionists would derive the properties of the remainder of the natural world, but which sometimes has the appearance of being a cornfield from which most of the crop has been harvested? The roster of the great discoveries of the past century accentuates present ennui: X-rays and radioactivity, electrons, cosmic rays, special relativity, quantum theory, general relativity, electron spin, mesons, fission, thermonuclear fusion, nucleogenesis, neutron stars, palaeomagnetism and plate tectonics, masers, lasers and transistors are each sufficient to make the shadowy top quark and Higgs boson seem minor events in such a calendar. A decade ago, there was a chance that the discovery of high-temperature superconductivity might have brought physics to life; it may yet, when somebody is able to account for it convincingly.

Matters inevitably seem worse because the comparison with parts of biology is unavoidable. What physicist would dare quarrel with the view that molecular genetics is where the action is and that, quite apart from the practical value of a listing of the human genome, the comparison between that and other genomes will surely provide such a detailed understanding of the intricacies of evolution that it will be possible to reconstruct the history of life on the Earth, even back to the beginning. Physicists cannot fail to have noticed the attention the newspapers pay to this prospect. Those who

remember the cheerful decades before 1970 will be doubly downcast.

In one sense, of course, physicists have brought these troubles on themselves. They have been ascetic to a fault. Faced with, say, the emergence of semiconductor devices, they have taught the essence of solid-state physics and then sent interested students off to departments of electrical or electronic engineering to make use of what they have learned. (It is a lucky accident that the scanning tunnelling microscope has made the structure of the surfaces of silicon and other metals, with or without absorbents, a part of physics.) Palaeomagnetism has similarly been hived off to geophysics, radio-astronomy to separate installations sometimes called observatories, much of fluid dynamics to geophysics (and the remainder to meteorology) and interesting questions such as nucleogenesis and interstellar chemistry to departments of astrophysics.

The reasons for this spawning of new disciplines from physics are plain and practical. They have to do with teaching. The hard core of the curriculum is demanding enough, from mechanics (classical and otherwise) and thermodynamics to optics and electromagnetism. And students must acquire that distinctive blend of knowing how to formulate problems accurately and of how to approximate the solutions of problems they cannot solve. The skill of designing an experiment that will actually measure what is intended is the essence of the craft. Stripped of the fun and games, it is bound to seem a little spartan, even dull. Maybe that is how Delhi students find it, but they are unlikely to be alone.

But none of this implies that physics is dead; it merely seems like that to some. The most casual glance through the specialist journals will show, for example, that the ingenuity now in evidence is more remarkable than it has ever been. The flurry of excitement, a few years ago, about the supposed 'fifth force' is a splendid proof that the traditions of elegant design are alive and well (see Fischbach, E. & Talmadge, C. *Nature* **356**, 207–215; 1992). The ingenuity lavished on the undeserving claims of cold fusion (see, for example, Williams, D. E. *et al. Nature* **342**, 375–384; 1989) is another proof of that. But, nowadays, the elegant experiments may consume years in preparation and execution. And then they are rarely as significant as, say, the Michelson–Morley experiment (for all the controversy its interpretation provoked).

Indeed, there now seems only a tenuous connection between human-scale laboratory experiments and the grand themes of physics — the structure of the Universe and the nature of the objects, made of particles, that it contains. (The hunts for bare quarks, sometimes with suitably instrumented domestic vacuum cleaners, and for magnetic monopoles, have been at least temporarily abandoned.) The small armies working around particle detectors are no longer on a human scale.

Even so, the experiments in the journals are not unconnected with the old grand themes. There is still, for example, much to be done to pin down quantum nonlocality (or to make actual the Einstein–Podolsky–Rosen *gedanken* experiment of 1935) and to verify that cyclic variations of some parameter characterizing a quantum system may change its quantum phase and thus its state. Even the meaning of the concept of a quantum state is now up for grabs experimentally, thanks to the ingenuity with which single atoms (or a few) can be physically trapped and then manipulated with lasers. It is no wonder that laser atomic physics is now a happy hunting-ground for experimentalists. It cannot be long before the transport of electrons through carriers of narrow dimensions (called 'quantum wires') follows suit.

But then, of course, there is a whole host of other gaps in understanding that cry out for attention. What about the quantization of the gravitational field, for example? In what sense can physics be said to be complete while such issues are wide open? If it comes to that, what is to be made of the puzzle over dark matter and the possibly connected difficulty about the Sun's underproduction of neutrinos? As always in physics, such gaps in understanding point to what are probably huge fields of intellectual activity still unexplored.

That is the sense in which it is especially misleading to think of the missing quark and the Higgs boson as the capstones that will bring physics to an end. The most likely outcome of the accelerator experiments being planned is either that these particles will not be found at all or that something else will appear instead. And then, of course, in the most constructive way, the field will come to life again. For a long time, the idea that there is a theory of everything will be forgotten. There will be no brownie points for wondering what it may be. Physics, in other words, is far from dead. **John Maddox**

Steps to unification

Richard Barvainis

THE extragalactic Universe contains a bewildering variety of objects — or so it seems. These include 'ordinary' galaxies, infrared galaxies, Seyfert and radio galaxies, BL Lac objects and quasars, the last coming in radio-loud and radio-quiet flavours. But perhaps some of these divisions are only apparent; what, if any, are the deeper connections between these seemingly diverse objects? Now Scoville and co-workers¹, reporting the discovery of carbon monoxide emission from the radio-loud quasar 3C48, may have provided another link in a growing chain of evidence for a unified picture of active galaxies.

Among the 'exotic' objects listed above, Seyfert galaxies have been known the longest. Named after the American astronomer Carl Seyfert (1911–1960), who in 1943 first recognized them as a class, they are spiral galaxies with very bright, point-like nuclei that radiate broad and narrow atomic emission lines, or sometimes only narrow lines, in the optical region. Those that emit both broad and narrow lines are classified as type 1; those that emit only narrow lines are type 2. Seyferts also emit radio waves, but only weakly, and are called 'radio-quiet'. The broad lines are thought to originate in clouds orbiting close to a very massive compact object, perhaps a black hole of ten million solar masses or more. The narrow lines arise considerably farther from the nucleus.

In 1985 came the discovery by Antonucci and Miller² that the type-2 Seyfert galaxy NGC1068 has broad lines which can be seen only in polarized (that is, reflected) light. From this followed the unified model for Seyferts, wherein all Seyfert nuclei are surrounded by an

opaque, dusty molecular torus. If viewed from above, the central broad line region can be seen directly, in addition to the narrow lines (which lie outside the torus). If viewed from the side, only the narrow lines are seen, unless some of the broad line radiation escaping from the top and bottom of the torus is reflected by ionized gas into our line of sight, as in the case of NGC1068 and, by now, a number of other Seyfert 2 galaxies. So the Seyfert type is determined solely by our viewing aspect relative to the torus.

At the time of this discovery, it was already becoming clear that Seyfert galaxies and quasars were closely related, and were perhaps just the same objects along a continuum of luminosity. The quasars, again, have broad and narrow lines, but are more luminous and farther away from us than the Seyferts, their nuclei being so bright that they outshine by large factors an entire galaxy of stars. Advances in imaging technology in the early 1980s led to the detection of galaxies surrounding many quasars, cementing the case in the minds of most astronomers that radio-quiet quasars are the high-luminosity analogues of Seyfert galaxies. A similar connection can be made between radio galaxies, which are similar to Seyferts but contain strong radio sources and appear to live in elliptical rather than spiral hosts, and radio-loud quasars³. But alas, the analogy was not quite perfect. The problem was that there appeared to be no 'quasar 2s', the luminous versions of Seyfert 2s.

Then along came the Infrared Astronomy Satellite (IRAS), which, in the mid-1980s, discovered hundreds of 'infrared galaxies', galaxies that emit most of their energy in the region between

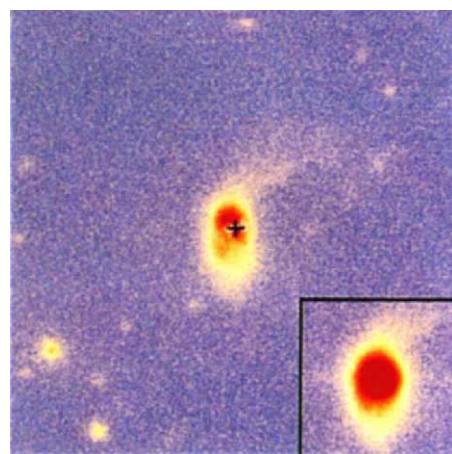


FIG. 2 Optical image of 3C48, showing the bright, point-like quasar nucleus plus the host galaxy (inset), and the host galaxy alone (main panel, quasar nucleus subtracted). The tail extending to upper right from the host galaxy was probably induced by tidal forces generated during an encounter with another galaxy. The cross in the quasar-subtracted image shows the position of the quasar nucleus; a second nucleus is apparent above and to the left of the cross, suggesting that two galaxies have recently merged to form what we see as the host of 3C48. (Image courtesy of A. Stockton.)

wavelengths of 1 μm and 1 mm. They are weak optically by comparison, but show strong emission from molecules, including carbon monoxide⁴. Many of them are as luminous as quasars, if all the energy emitted across the spectrum is added up, but they show no direct evidence of quasar-like activity such as broad emission lines. Nevertheless, on the basis of their luminosities and the discovery that their space density is comparable to that of quasars, it was hypothesized that they were related in an evolutionary sense^{5,6} to what are collectively known as active galactic nuclei, or AGNs — quasars, Seyfert galaxies and radio galaxies. Could they be the quasar 2 objects?

To make the connection, we need to understand what triggers a quasar to light up so brightly, and how it is fuelled. The current picture remains speculative, but seems intuitively likely. The most important single fact is that the ultraluminous infrared galaxies almost always show signs of interaction or collision with another galaxy. Nearby companions, distorted morphologies, tidally induced tails and double nuclei all suggest current or recent mergers or interactions⁷. It turns out that AGNs also show a significant excess of nearby companions and other interaction signposts. If we assume that supermassive black holes exist in galactic nuclei (for which there is evidence in nearby galaxies as well as in AGNs⁸), and that a quasar is born and incandesces as fuel (gaseous matter) is heated while falling into the hole, then perhaps we can see why a galactic collision could be the trigger.

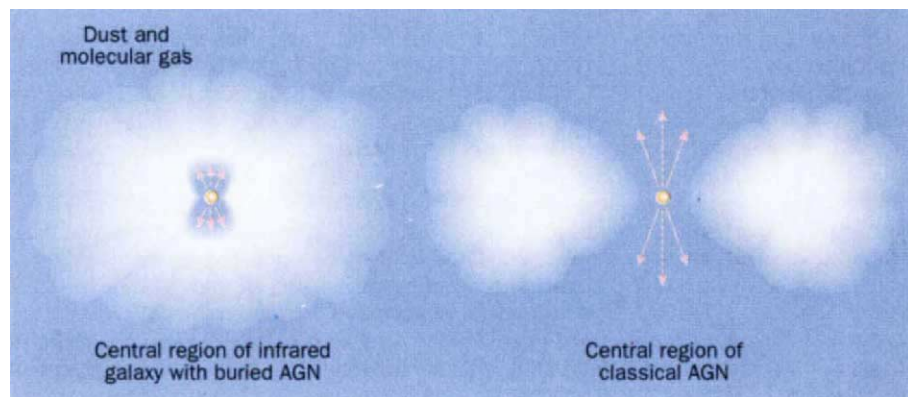


FIG. 1 Hypothetical evolutionary sequence. Left, cutaway view of central region of infrared galaxy shortly after merger. A cloud of dusty gas surrounds the reactivated active galactic nucleus (AGN), which has begun to clear out a cavity. Right, at later times, dust and gas have been partially removed, revealing the AGN along polar directions. An opaque torus of infalling material remains in the equatorial plane. Seen from above, this would be a 'type 1' Seyfert galaxy; seen from the side, it would be classified as 'type 2'.

Collisions stir up the diffuse matter in a galaxy, perhaps driving large clouds of dust and gas into the nucleus and thereby feeding the black hole.

The evolutionary sequence is thought to go something like this. Gas stirred up or deposited during a merger feeds the black hole, but at first the resultant quasar is deeply buried in a shroud of gas and dust and cannot be seen directly. Its light at this stage will be absorbed and reprocessed by the dust, which emits in the infrared and thereby results in an infrared galaxy. Eventually the gas and dust are blown away within some cone by the quasar light, but continue to fall inwards in the orthogonal plane, forming the torus (see Fig. 1). At this point, we have a classical AGN. Partially cleared nuclei might represent the quasar 2 stage.

How, then, does Scoville and co-workers' detection of carbon monoxide in 3C48, a radio-loud quasar, add to this picture? Molecular gas has already been detected in several radio-quiet quasars, which are thought to be located in gas-rich spiral galaxies⁹⁻¹¹. Recalling that radio-loud quasars appear to live in elliptical galaxies, which are normally gas- and dust-poor, it is at first sight surprising that copious amounts of molecular gas should be found there. But the galaxies hosting radio-loud quasars have been found to be unusually blue for ellipticals, indicating a young, newly formed stellar population rather than the normal old population. This would seem to require the presence of large amounts of gas and dust to form the stars. 3C48's host galaxy is quite blue¹², and it is also a strong infrared source, having been detected by the IRAS satellite. This then suggests that 3C48 has just emerged from a formative cocoon of gas and dust⁶, as outlined above, with the additional proviso that it is undergoing a burst of star formation (also fairly common in infrared galaxies). Support comes from optical images showing clear signs of a recent merger (see Fig. 2).

So here we have evidence for the possible evolutionary link between gas-rich infrared galaxies, and quasars residing in ordinarily gas-poor elliptical hosts. Spir-

als, on the other hand, are gas-rich, and so fit more naturally into the scheme. With the detection of molecular gas in 3C48, radio-loud quasars, and their elliptical hosts, can now be embraced more fully into this aspect of unification.

But where does all the gas and dust in 3C48 come from, if ellipticals are normally so gas-poor? Recent theories of galactic cannibalism suggest that giant elliptical galaxies may form and grow by 'eating'

many hapless spirals over the course of billions of years¹³. Perhaps with each meal, rich in dust and gas, the black hole in the centre is reactivated, emitting the tremendous burst of radiant energy we call a quasar. □

Richard Barvainis is at the Haystack Observatory, Massachusetts Institute of Technology, Westford, Massachusetts 01886, USA.

TECTONICS

Converging more slowly

Robert S. Yeats

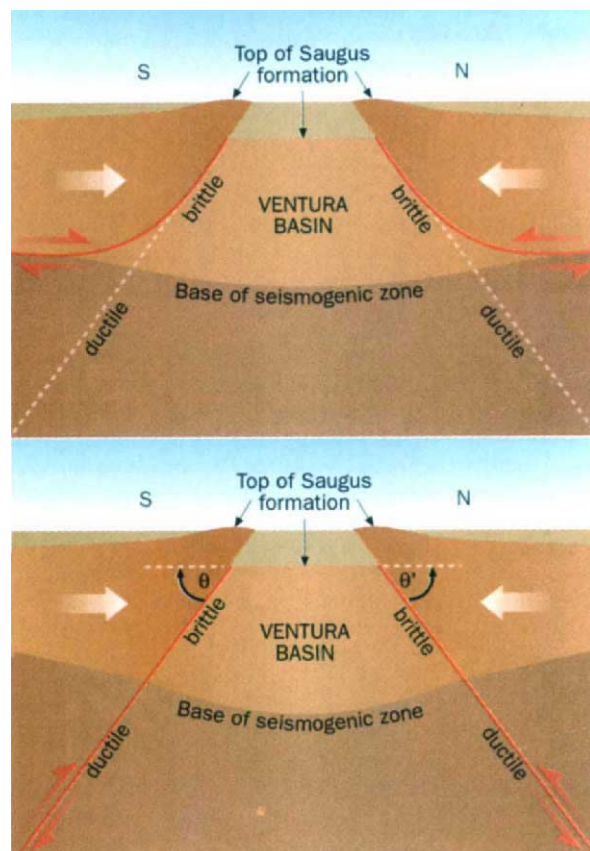
SPACE geodesy, principally done by means of the Global Positioning System satellites and very-long-baseline interferometry, demonstrates that the relative rates of motion of the Earth's plates, based on geological timescales measured over 10^5 – 10^6 years, are confirmed by motions based on timescales measured in years or in decades¹. Now the challenge is to analyse regions of less than plate dimensions where the displacement rates using the two timescales do not agree. Are differences due to higher-frequency rate changes, errors in measurements or a misunderstanding of the tectonics?

One such region is the Ventura basin in California, where Donnellan, Hager and King², on page 333 of this issue, document

a shortening rate of 7 – 10 mm yr⁻¹ over a period of 4.6 years. This rate agrees with a shortening rate of around 6 mm yr⁻¹ in the eastern Santa Barbara Channel^{3,4}, but both rates are lower by at least a factor of 2 than geologically based shortening rates averaged over the past several hundred thousand years⁵⁻⁷.

The geologically based rates assume that slip on moderately dipping faults (the San Cayetano and Red Mountain Faults on the north and Oak Ridge Fault on the south) transfers at depth to horizontal shortening at the base of brittle crust (see figure). These regions of horizontal slip-page, or décollements, converge on the Ventura basin^{4,7}. The décollement model was inspired by the discovery that upper-

Alternative tectonic models for the Ventura basin, California. Convergence (open arrows) is accommodated by reverse faulting on the margins of the basin. If the reverse faults (shown in red) flatten to décollements (top), as suggested by earthquake focal mechanism solutions, then the convergence rate is equal to the sum of the slip rates v and v' on the faults as measured near the surface. If the reverse faults continue into the lower crust as ductile shear zones (bottom), then the convergence rate is reduced to $v \cos \theta + v' \cos \theta'$ in better agreement with rates derived from space geodesy (Donnellan, Hager and King²).



1. Scoville, N. Z., Padin, S., Sanders, D. B., Soifer, B. T. & Yun, M. S. *Astrophys. J.* **415**, L75–L77 (1993).
2. Antonucci, R. & Miller, J. *Astrophys. J.* **297**, 621–632 (1985).
3. Barthel, P. *Astrophys. J.* **336**, 606–611 (1989).
4. Sanders, D. B., Scoville, N. Z. & Soifer, B. T. *Astrophys. J.* **370**, 158–171 (1991).
5. Antonucci, R. J. & Olszewski, E. W. *Astr. J.* **90**, 2203–2206 (1985).
6. Sanders, D. B. *et al. Astrophys. J.* **325**, 74–91 (1988).
7. Stockton, A. & Ridgeway, S. E. *Astr. J.* **102**, 488–494 (1991).
8. Kormendy, J. *Astrophys. J.* **325**, 128–141 (1988).
9. Sanders, D. B., Scoville, N. Z. & Soifer, B. T. *Astrophys. J.* **335**, L1–L4 (1988).
10. Barvainis, R., Alloin, D. & Antonucci, R. *Astrophys. J.* **337**, L69–L72 (1989).
11. Alloin, D., Barvainis, R., Gordon, M. A. & Antonucci, R. *Astr. Astrophys.* **265**, 429–436 (1992).
12. Boroson, T. B. & Oke, J. B. *Astrophys. J.* **281**, 535–544 (1984).
13. Kormendy, J. & Sanders, D. B. *Astrophys. J.* **390**, L53–L56 (1992).

Good behaviour in conservation

THE difficulties of reintroducing captive-bred species into the wild are epitomized by attempts to conserve the waldrapp (or bald) ibis, *Geronticus eremita*, shown here. There are now fewer than 300 birds left in the wild, almost all living in a single reserve in Morocco, whereas in the sixteenth and seventeenth centuries the species was found in the Middle East and North Africa, and as far north as Switzerland, Austria and Germany. The last wild Turkish population became extinct in 1989 as a result of pesticide pollution and habitat disturbance.

Ironically, although numbers have plummeted in the wild, the bird flourishes in captivity and about 800 captive individuals exist. The waldrapp ibis is in many ways a prime candidate for reintroduction. It appears to be a prolific breeder in the right conditions, and it seems to have disappeared from Europe largely as a result of human persecution: the eggs and chicks were considered a great delicacy.

There have been two attempts at reintroducing waldrapps, one in Turkey in 1981, and another in Israel two years later. Both seem to have failed. Two hundred birds were released, but immediately dispersed and none has survived.

The failures were thought to be due to a breakdown in the normal social behaviour of the ibises. Waldrapp ibises are social birds with extended parental care, and it seems that much of this behaviour must be learnt, but is not in captivity. Can it be taught to captive-bred individuals? It appears that it can, as K. Pegoraro and colleagues from the University of Innsbruck told an ethological conference this

autumn*. Young waldrapp ibises were hand-reared in Austria by human foster parents who taught them to find their way to nearby fields to forage. They also had to be taught to recognize predators, human and otherwise, and other dangers such as cars. Waldrapp ibises have no alarm call and so were taught to recognize that of a chough raised with them, which was easily mimicked by their human guardians when danger threatened. Even mutual preening had to be taught.

Eventually their behaviour was very simi-

*XXXIII International Ethological Conference, Torremolinos, Spain, 1–9 September 1993.

lar to that observed in the wild population in Morocco, although it is not known how successfully these captive-bred birds will breed. There are no plans to establish a permanent colony in Austria. But birds hand-reared according to similar principles may be released in the National Park Cabo de Gata-Níjar in Spain, where the climate, topography and availability of prey are very similar to conditions at the reserve in Morocco.

W. J. Sutherland

William J. Sutherland is in the School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

mantle structure beneath the western Transverse Ranges does not match the crustal structure⁸, so it looks as if the fault is diverted in the ductile layer between brittle crust and mantle. Low-angle thrust focal-mechanism solutions are found at mid-crustal depths, just where the décollement should occur⁹. A geological precedent may be found in the crystalline thrust sheets or tectonic 'flakes' in the eastern Alps¹⁰ and the San Gabriel Mountains of southern California¹¹. These thrusts, now at the surface, formed at depths and temperatures assumed to be near the base of brittle crust.

The shortening rates based on the décollement model are at the high end of those rates possible based on known slip rates on other faults in southern California¹². The discrepancy with the rates based on tectonic geodesy can be resolved if, first, the reverse faults do not flatten to décollements but continue downward into the lower crust as ductile shear zones¹³, and second, the age estimate of the top of the Saugus formation, on which the geological rates are based, is too young.

If the faults do not flatten to zero dip at

a décollement, the horizontal convergence rate would be the fault slip rate times the cosine of the fault dip angle in the lower crust (see figure). This would reduce the convergence rates to between 70 and 40 per cent of their values based on the décollement model. The age of the top of the Saugus is 200 kyr, if we calculate it from estimates of the racemization rate of amino-acids¹⁴. Because of uncertainties in the kinematics of amino-acid racemization and the effects of other factors, notably temperature, these estimates could be wrong. The geodetic results support an older age; let us take 500 kyr, based on a magnetostratigraphic section in the east Ventura basin¹⁵ as representative for the Ventura basin as a whole. With this revised age of the top of the Saugus formation and the assumption that reverse faults continue into the lower crust, the convergence rates based on geology are 7–9 mm yr⁻¹, in agreement with the geodetic results. These rates are about the same as those for the preceding 500 kyr, whereas the younger age of the Saugus required an abrupt increase in convergence rate since the Saugus was deposited.

Results from the Global Positioning

System suggest that the southern part of the survey area is rotating, and this points to a weakness in the balanced cross-section method on which the geological estimates of convergence rates are based. Balanced cross-sections of deformed regions are constructed so that they can be restored to their undeformed state without loss of cross-sectional area. This is based on the assumption that deformation

- Gordon, R. S. & Stein, S. *Science* **256**, 333–342 (1992).
- Donnellan, A., Hager, B. H. & King, R. W. *Nature* **366**, 333–336 (1993).
- Larson, K. M. & Webb, F. H. *Geophys. Res. Lett.* **19**, 1491–1494 (1992).
- Larsen, S. C., Agnew, D. C. & Hager, B. H. *J. geophys. Res.* **98**, 2119–2133 (1993).
- Yeats, R. S. *J. geophys. Res.* **88**, 569–583 (1983); **93**, 12137–12149 (1988).
- Rockwell, T. K. *Geol. Soc. Am. Bull.* **100**, 500–513 (1988).
- Huflile, G. J. thesis, Oregon State Univ. (1992).
- Hadley, D. & Kanamori, H. *Geol. Soc. Am. Bull.* **90**, 1449–1457 (1978).
- Webb, T. H. & Kanamori, H. *Bull. seism. Soc. Am.* **75**, 737–757 (1985).
- Oxburgh, E. R. *Nature* **239**, 202–204 (1972).
- Yeats, R. S. *Geology* **9**, 16–20 (1981).
- Weidon, R. & Humphrey, E. *Tectonics* **5**, 33–48 (1986).
- Yeats, R. S. & Huflile, G. J. *Eos* **73**, 590 (1992).
- Yerkes, R. F., Sarna-Wojcicki, A. M. & Lajoie, K. R. *U.S. Geol. Surv. Prof. Pap.* **1339**, 169–178 (1987).
- Levi, S. & Yeats, R. S. *Tectonics* **12**, 688–702 (1993).
- Jackson, J. & Molnar, P. *J. geophys. Res.* **95**, 22073–22087 (1990).

occurs in the line of section, and that material does not move into or out of the section. The assumption is clearly untrue if rotation at right angles to the line of the cross-section has occurred within the time frame of interest. Evidence for tectonic rotation confirms observations made elsewhere in the western Transverse Ranges that were also based in part on tectonic space geodesy¹⁶.

The work of Donnellan, Hager and King does not merely confirm tectonic rates based on timescales of 10^5 – 10^6 years, it revises these rates, and it requires a re-evaluation of tectonic models on which

these rates are based. At present, satellite observations span too short a time to allow comparison with longer timescales where rates are slow. Longer observation periods are also required to confirm that these rates are representative of the long term, rather than being a transient state related to locking of faults before a major earthquake. □

Robert S. Yeats is in the Department of Geosciences, Oregon State University, Corvallis, Oregon 97331, and the Southern California Earthquake Center, Los Angeles, California 90089, USA.

DNA REPAIR

The nick in time

Bernard S. Strauss

RESEARCH into nucleotide excision-repair in yeast is in ferment, the latest manifestation of that happy state of affairs coming in the form of the paper by Habraken and colleagues on page 365 of this issue¹. The authors show that one of seven yeast *RAD* genes required to incise ultraviolet-damaged DNA, *RAD2*, encodes an endonuclease specific for single-stranded DNA. This discovery closely follows reports that two other yeast genes essential for the repair of damaged DNA, *RAD1* and *RAD10*, together encode a second DNA endonuclease^{2,3}.

These are solid biochemical and genetic accomplishments. To understand why they should be of general interest, one needs to know that, 40 years ago, one of the arguments used to justify the idea that DNA was genetic material was its apparent metabolic stability⁴. Whereas RNA and protein 'turned over', there seemed to be convincing evidence that atoms were permanently incorporated into DNA. The true state of affairs has turned out to be practically the reverse — organisms continually monitor the integrity of their genetic material, replacing those portions that have been damaged by either exogenous or endogenous events. The efficiency with which these processes occur masks their action and contributes to the view that repair is an extraordinary biological event. In fact, hydrolysis of purine bases alone results in the turnover of as many as 10,000 nucleotides each day in a human cell⁵.

The overall sequence of events in one of the main pathways of repair has been known for at least 30 years^{6–8}. Broadly, this idea of an initial recognition of DNA damage, followed by two incisions flanking the damage on one strand, removal of the damaged section and its replacement by synthesis using the undamaged strand as a template, is the 'cut and patch' mechanism found in college textbooks.

This simple and satisfying scheme, along with the perfectly normal behaviour of ultraviolet-sensitive mutants of the much studied *Escherichia coli*, except when challenged by ultraviolet light or other DNA-damaging agents, seemed to indicate a peripheral role for repair in general biology.

Eukaryotes are different, however. They combine excision-repair functions with other vital cell activities, and the finding that components of the excision-repair system are also major players in the initiation of transcription was described in News and Views in May⁹. Here it needs to be restated that one reason for the level of activity in this field is that there is remarkable homology between the genes involved in excision repair in all eukaryotes. *RAD2*, for instance, is the homologue of the human xeroderma pigmentosum (XP) group G complementing gene. *RAD3* and its human homologue *XPD* both encode DNA helicases¹⁰. The *RAD25* protein and its human counterpart, *XPB*, both have helicase motifs, and *XPB* corresponds to part of the transcription factor complex TFIIF (ref. 11).

Excision repair can be accomplished in reconstructed systems of *E. coli* proteins, but until recently eukaryotic repair has evaded biochemical analysis *in vitro*. The difficulties have been at least partially overcome and systems have become available for studies with human cell extracts^{12,13} and yeast¹⁴. Earlier this year, Sancar and his group showed¹⁵ that one of the main photoproducts, the cyclobutane thymine dimer, was excised as part of a 29 nucleotide oligomer. But it was not determined whether the two incisions required were made simultaneously, as in *E. coli*, or sequentially.

Habraken and colleagues now resolve this question. *RAD2* and *RAD1* plus *RAD10* gene products are required for incision¹, *RAD1* and *RAD10* acting

together as a complex^{2,3}. Both *RAD2* and the *RAD1/RAD10* complex are endonucleases that cleave single-stranded DNA. *RAD1/RAD10* also makes single-stranded nicks in double-stranded DNA, and the nicking occurs more readily as negative supercoiling increases². *RAD2* does not nick supercoiled double-stranded DNA¹. Incision probably occurs sequentially, with the first nick being made in supercoiled double-stranded DNA by *RAD1/RAD10*, and the second in unwound single-stranded DNA by *RAD2*.

Preferential incision of damaged DNA requires not only *RAD1/RAD10* and *RAD2*, but also *RAD3*, *RAD4*, *RAD14* and *RAD25*. What might these other proteins be doing? Neither *RAD2* nor *RAD1/RAD10* recognize damaged DNA^{1,2}, and Habraken *et al.* suggest that, subsequent to damage recognition by *RAD14*, the helicase activity of *RAD3* and *RAD25* may convert the damaged DNA into a substrate for *RAD1/RAD10* which makes the first nick. Further helicase action then produces a single strand which can be incised by *RAD2*.

The nucleotide excision-repair system recognizes a wide range of DNA damage, but it remains to be seen how it does so. Even more intriguing is the question of the stability of the complexes involved in transcription and in repair. Are all the proteins continually bound together or are the complexes dynamic, with a central core and with additional proteins being added as needed for particular functions? And does the finding that the *RAD1/RAD10* complex is involved in mitotic recombination stimulated by transcription² mean that transcription/repair complexes are permanently associated? □

Bernard S. Strauss is in the Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago 60637, USA.

- Habraken, Y., Sung, P., Prakash, L. & Prakash, S. *Nature* **366**, 365–368 (1993).
- Sung, P., Reynolds, P., Prakash, L. & Prakash, S. *J. biol. Chem.* **268**, 26391–26399 (1993).
- Tomkinson, A., Bardwell, L., Tappe, N. & Friedberg, E. *Nature* **362**, 860–862 (1993).
- Mazia, D. in *Modern Trends in Physiology and Biochemistry* (ed. Barron, E.) 77–122 (Academic, New York, 1952).
- Lindahl, T. *Nature* **362**, 709–715 (1993).
- Setlow, R. & Carrier, W. *Proc. natn. Acad. Sci. U.S.A.* **51**, 226–231 (1964).
- Boyce, R. & Howard-Flanders, P. *Proc. natn. Acad. Sci. U.S.A.* **51**, 293–300 (1964).
- Pettijohn, D. & Hanawalt, P. *J. molec. Biol.* **9**, 395–410 (1964).
- Bootsma, D. & Hoeijmakers, J. *Nature* **363**, 114–115 (1993).
- Sung, P. *et al.* *Nature* **365**, 852–855 (1993).
- Schaeffer, L. *et al.* *Science* **260**, 58–63 (1993).
- Wood, R., Robins, P. & Lindahl, T. *Cell* **53**, 97–106 (1988).
- Singh, Ullah, Husain, I., Carlton, W. & Sancar, A. *Nucleic Acids Res.* **17**, 4471–4484 (1989).
- Wang, Z., Wu, X. & Friedberg, E. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4907–4911 (1993).
- Huang, J., Svoboda, D., Reardon, J. & Sancar, A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3664–3668 (1992).

Silence of the genes

M. Azim Surani

DURING development of the embryo and after birth, some genes can 'remember' which parent they came from. This remarkable feat is possible because they inherit molecular imprints from the respective germ lines. As a result, only one parental copy is active in developing embryos. Imprinting starts in the germ line, so a heritable molecular tag is needed; this tag should be able to regulate gene expression and be removable during each cycle of germ cell development. DNA methylation may be one of the molecular markers, and the fate of parental imprints in mutant mice with impaired DNA methylation is now reported on page 362 of this issue by Li, Beard and Jaenisch¹. The outcome is striking — the imprints on three genes (*Igf-2r*, *Igf-2* and *H19*) are lost, resulting in the expression or repression of both parental copies and the death of mutant embryos.

To begin at the end, methylation differences between parental copies of the three imprinted genes are detected in post-implantation embryos (*a* in the figure). The gene encoding the receptor for insulin-like growth factor *Igf-2* (*Igf-2r*) is an imprinted gene with an active maternal copy and an inactive paternal copy². In this gene, two distinct CpG islands show parental-origin-specific methylation differences: the intronic CpG island 27 kilobases downstream of the promoter is methylated on the maternal copy whereas the promoter region is methylated on the inactive paternal gene³. The *Igf-2* and *H19* genes are closely linked but are imprinted reciprocally, in that the paternal copy of the *Igf-2* (ref. 4) and the maternal copy of *H19* (ref. 5) are expressed during development. Two regions in this domain have so far been shown to be differentially methylated. Sites 3 kilobases upstream of *Igf-2* (ref. 6) and the promoter CpG island of the *H19* gene are both methylated on the paternal chromosome, although the difference in the former is less pronounced⁶⁻⁸.

Selection

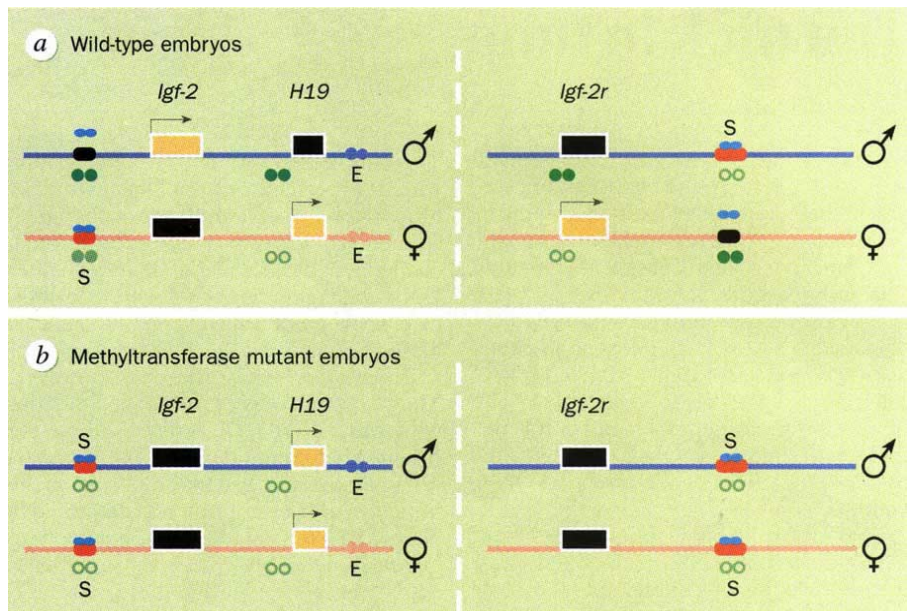
Can these differences in methylation account for the selection of the active parental gene? CpG islands are predominantly associated with promoters and it is rare for them to be methylated. When methylation does occur, for example with the X-linked genes, this results in gene silencing⁹. The *Igf-2r* gene is unusual in having an intronic CpG island that is methylated on the active maternal copy of the gene. It may be that here the intronic island is acting as a powerful silencer on the paternal chromosome when it is unmethylated, perhaps by bind-

ing repressor molecules³. If so, then the silencer must function quite early to repress the gene because the promoter is methylated subsequently and relatively late in development³.

With *Igf-2* and *H19*, things are more complex. First the 3-kilobase region upstream of the *Igf-2* gene is less methylated on the maternal chromosome and could act as a silencer on this chromosome, as has been suggested for the *Igf-2r* gene⁶. Second, the observed methylation and

90-kilobase intergenic region that could harbour other elements, for example regulators of the three *Igf-2* promoters.

Mammals cannot survive without maintaining strategic DNA methylation, to judge by the death *in utero* of mice homozygous for a DNA methyltransferase mutation¹⁰. These mutant embryos also show a loss of methylation imprints and aberrant expression of the three genes already discussed (*b* in the figure)¹. In the *Igf-2r* gene, the intronic CpG region on the maternal chromosome becomes demethylated and is presumably able to act as the postulated silencer. Indeed, both parental chromosomes now become functionally paternal and no *Igf-2r* is



Representation of, *a*, wild-type embryos showing methylated (green circles) and unmethylated (light green or open circles) regions on parental chromosomes with the three imprinted genes. S, the putative silencer element that is active when unmethylated (red), following binding of repressor molecules (blue). When methylated, these elements are not active (black). Yellow boxes with arrows show active copies of genes and the direction of transcription; black boxes denote silent copies of the parental genes. E, enhancers. *b*, Depiction of changes in methylation and expression of imprinted genes in methyltransferase mutant embryos associated with the loss of methylation imprints. The upstream region of *Igf-2* was not examined in the mutant embryos, and possible changes in this region are indicated.

condensed chromatin of the inactive paternal *H19* promoter may be the controlling event^{7,8}, so that imprinting of *H19* and *Igf-2* might be linked mechanistically if the *Igf-2* and *H19* promoters compete for the two enhancers downstream of *H19* (refs 6, 8). The enhancers may activate *Igf-2* expression, but only if the *H19* gene is methylated and inactive, as on the paternal chromosome⁸. Interestingly, there are no detectable differences in DNA methylation or chromatin structure between the *Igf-2* parental genes, and the inactive maternal *Igf-2* gene retains an open chromatin configuration and low-level transcription⁶. As yet there is no evidence to distinguish between the competition and silencing models, although they are by no means mutually exclusive. Furthermore, little is known about the

detectable in mutant embryos (albeit only in the presence of the stronger methyltransferase mutant allele *MTase*^s; see ref. 1).

For the *Igf-2/H19* domain, Li *et al.* show that the paternal *H19* gene becomes demethylated and that both parental genes are active in mutant embryos. They did not examine the methylation status of the upstream *Igf-2* region, but it is probably demethylated and the paternal *Igf-2/H19* domain becomes functionally maternal. Indeed, there is a striking absence of *Igf-2* transcripts in the mutant embryos, which would be expected if the upstream region of *Igf-2* acts as a silencer when the region is unmethylated. The prediction would be the same if, with activation of both *H19* genes, the enhancers could not interact with the *Igf-2* promoters, resulting in loss

of *Igf-2* transcription^{1,8}. Intriguingly, unlike the loss of *Igf-2r* activity, *Igf-2* activity is lost with the weaker *MTase*ⁿ mutant allele, suggesting that there is a greater sensitivity of the *Igf-2* upstream region and the *H19* gene to methylation changes. Further investigation should resolve whether or not there is a direct relationship between the imprinting of the *H19* and *Igf-2* genes.

Demethylation

Another question — that of whether differential methylation patterns are inherited from the germ line — is pertinent because such imprints must survive the striking genome-wide demethylation that accompanies pre-implantation development^{11–13}. Overall DNA methylation is exceedingly low in blastocysts, which may be crucial for establishing pluripotential epiblast cells. It might be informative to test whether low-level transcription of imprinted genes occurs in blastocysts and, if it does, to see whether both parental alleles of imprinted genes are expressed. Later, with the onset of gastrulation, *de novo* methylation occurs around day 6.5 of development and includes methylation of the paternal *H19* promoter^{7,14}. Whether the primordial germ cells escape that methylation is a moot point, but they have the capacity to erase methylation and establish the new germline pattern^{15,16}.

The *Igf-2r* intronic region is perhaps the most clear-cut example of the inheritance of methylated sites from oocytes that remain intact in blastocysts^{3,14}. Some sites in the upstream region of the *Igf-2* gene are also methylated, but in both sperm and oocytes; although many of these sites undergo demethylation in pre-implantation embryos, one of the paternally inherited sites remains methylated at least in some of the cells¹⁴. In the *H19* gene, no sites have yet been identified that are inherited as differentially methylated from gametes. At present it is not possible

to say unequivocally if DNA methylation is the signal that specifies the germline imprint for any or all imprinted genes. But the inability of methyltransferase mutant embryos to maintain normal regulation of imprinted genes confirms the view that DNA methylation is a key event in the imprinting process.

Of the three imprinted genes, the function of *H19* gene, which appears not to encode a protein, remains enigmatic, although recent evidence indicates that the *H19* RNA may have tumour-suppressor activity¹⁷. We already know that imbalanced expression of imprinted genes is lethal. Embryos with the maternal disomy for chromosome 7 have excess *H19* and no *Igf-2* (just like the methyltransferase mutant embryos), and *Tme*

mutation embryos lack the functional *Igf-2r* gene; both are lethal^{2,7}. This means that the methyltransferase mutants cannot survive because of the perturbation of the three genes. But there may be more widespread disturbance of gene regulation, and these mutant mice could prove useful for investigating the extent to which DNA methylation controls early development. More notably, though, it should now also be feasible to investigate the influence of DNA methylation in the germ line and the consequences for genomic imprinting. □

M. Azim Surani is at the Wellcome/CRC Institute of Cancer and Developmental Biology, Tennis Court Road, Cambridge CB2 1QR, UK.

MATERIALS SCIENCE

Confusion by design

A. Lindsay Greer

GLASSY metals mostly teeter on the verge of viability. They are difficult to produce in the first place because they require such high cooling rates, and they show an ungrateful tendency to undo the good work by crystallizing on heating, hindering the customary hot-forming processes and limiting applications. This has been frustrating, to put it mildly, for those wishing to exploit the unusual properties of the materials, so last month's news of the work of Peker and Johnson¹ should come as music to their ears. By including so wide a range of metallic elements that the alloy was essentially too confused to crystallize, they have devised an exceptionally stable, user-friendly glass.

Since the first reports of metallic glasses produced from the liquid², their formation has been associated with ultra-high cooling rates, typically over 10⁵ kelvin per second, and this has restricted production to forms with very thin cross-section (ribbon, wire or droplets). This isn't an inescapable requirement: there have been a few notable exceptions, for instance Pd₄₀Ni₄₀P₂₀ glass (all compositions are in atomic per cent), for which bulk sections 10 millimetres across could be produced at a cooling rate of only 1 kelvin per second — the catch being that the alloy must be kept very clean in order to inhibit heterogeneous nucleation³.

The field really took off, though, with the demonstration that excellent glass-forming ability could be found even without noble metals (which are costly) or metalloids (whose inclusion somewhat detracts from the idea of having a fully metallic glass, and may lead to brittleness). The group of Inoue and Masumoto⁴ showed that alloys of the family Ln–Al–TM (where Ln represents a lanthanoid

and TM a transition metal; an example is La₅₅Al₂₅Ni₂₀) could be cast to form cylinders and sheets a few millimetres in cross-section. The group went on to show that it was not essential for a lanthanide to be the base element (for example, they obtained good results from Mg–TM–Ln alloys such as Mg₆₅Cu₂₅Y₁₀)⁵, or indeed to be present at all (as in Zr–Al–TM alloys)⁶.

What was going on? In the latter alloy type⁶, the Japanese group showed that the best results could be obtained with more than three elements, stirring in, for example, two different transition metals (Zr₆₅Al_{7.5}Ni₁₀Cu_{17.5}). Considering variations in atomic radius, they suggested that a mixture of elements leads to dense packing in the liquid state, and the resulting stability favours glass formation over crystallization.

Peker and Johnson have now taken the logical next step, with the addition of beryllium. Their composition — Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10.0}Be_{22.5} — sets the pattern for anomalously easy glass formation in alloys, and its complexity is just the key to its success.

It seems to have been widely accepted by those working on metallic glass formation that a 'confusion principle' should apply: the more elements involved, the lower the chance that the alloy can select viable crystal structures, and the greater the chance of glass formation. So why have there been so few attempts to exploit the idea? The answer, it seems, is that 'two's company, three's a crowd' to those wishing to analyse fully the thermodynamics and kinetics of alloy transformations — the more elements, the more difficult the analysis.

The recent results show that the principle, when applied with care, works very

- Li, E., Beard, C. & Jaenisch, R. *Nature* **366**, 362–365 (1993).
- Barlow, D. P., Stöger, R., Herrmann, B. G., Saito, K. & Schweifer, N. *Nature* **349**, 84–87 (1991).
- Stöger, R. *et al.* *Cell* **73**, 61–71 (1993).
- De Chiara, T. M., Robertson, E. J. & Efstratiadis, A. *Cell* **64**, 849–859 (1991).
- Bartolomei, M. S., Zemel, S. & Tilghman, S. M. *Nature* **351**, 153–155 (1991).
- Sasaki, H. *et al.* *Genes Dev.* **5**, 1843–1856 (1992).
- Ferguson-Smith, A. C., Sasaki, H., Cattaneach, B. M. & Surani, M. A. *Nature* **362**, 751–754 (1993).
- Bartolomei, M. S., Webber, A. L., Brunkow, M. E. & Tilghman, S. M. *Genes Dev.* **7**, 1663–1673 (1993).
- Bird, A. P. *Nature* **321**, 209–213 (1986).
- Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915–926 (1992).
- Monk, M., Boubelik, M. & Lehnert, S. *Development* **99**, 371–382 (1987).
- Howlett, S. K. & Reik, W. *Development* **113**, 119–127 (1991).
- Kafri, T. *et al.* *Genes Dev.* **6**, 705–714 (1992).
- Brandeis, M. *et al.* *EMBO J.* **12**, 3669–3677 (1993).
- Chaillet, J. R., Vogt, T. F., Beier, D. R. & Leder, P. *Cell* **66**, 77–83 (1991).
- Ueda, T. *et al.* *Development* **116**, 831–839 (1992).
- Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E. & Tycko, B. *Nature* **365**, 764–767 (1993).

effectively. The most 'confusing' elements will be those that differ most from each other in size, and here lies the key to the use of beryllium: it is the smallest metallic atom. Its addition should offer extra scope for the design of other glass-forming compositions, allowing elements to be selected for properties such as magnetism and not just for their glass-forming ability.

The magnitude of the advance for metallic glasses can be gauged most directly from sample size: a conventional alloy made glassy by ultra-rapid quenching is limited to a section thickness of less than 0.1 millimetres, whereas Peker and Johnson, using simple water quenching (cooling at 10 kelvin per second or less), have made fully glassy rods 14 millimetres in diameter, and this is not regarded as an upper limit. With samples of such a size, whole new areas of application become available to amorphous alloys, allowing us to exploit desirable properties such as high strength, corrosion resistance and soft magnetism.

Peker and Johnson have already shown that it is possible to cast and machine small components. The unusual resistance to crystallization would make possible the shaping of sheet metal by viscous flow at elevated temperature. This perfect superplasticity is widely exploited in the forming of conventional oxide glasses, but would be a considerable novelty for metallic alloys.

Besides being of interest for large samples, the formability of the new glassy alloys may be useful for very small components. Microforming of gear wheels with feature resolution of 1 micrometre has been reported⁷, the advantage of the glass being its homogeneous isotropic deformation. And a bonus is that the new materials can be consolidated without crystallizing; making sizeable components from the flimsy ribbons or droplets of the old method was an art in itself.

With careful design of compositions, easy glass formation should be possible in quite a variety of multicomponent alloy systems. Given the sheer complexity of analysis and calculation of phase diagrams and kinetics for multicomponent systems such as these, it seems likely that metallic glasses have a few surprises still in store. □

A. Lindsay Greer is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

Hot lasers, cool molecules

Herschel Rabitz

HOT molecules may be generated easily as the result of chemical reactions or through external heating. Cooling such molecules (reducing their vibrational temperature) allows a detailed look at molecular dynamics and might ultimately help chemists to realize their dream of controlling reaction pathways at this smallest of scales. Bartana and co-workers have come up with a scheme for using finely controlled and designed laser pulses to do just that (*J. chem. Phys.* **99**, 196–210; 1993).

The authors consider an ensemble of molecules vibrationally excited in their ground electronic state. A traditional approach to molecular cooling might be to remove the population amongst the most vibrationally excited levels and transfer that population elsewhere, perhaps to an excited electronic state. In contrast, Bartana and colleagues consider a novel cooling procedure which does not change the overall population in the ground state, but rather rearranges it through coherent radiative coupling with an excited electronic state.

The best analogy is with an everyday refrigerator, which employs a working fluid to carry heat away from the cold region (the refrigerator interior) and dump it into a waste heat sink (the surrounding kitchen). The molecular-scale quantum mechanical refrigerator operates under the same principles, but now the desired cold 'region' is the ground electronic state of the molecule, the working fluid is a coherently designed optical field, and the heat sink is an excited electronic state.

Population

A carefully designed optical field with coherent phase structure can, say the authors, maintain the overall population in the ground electronic state, while rearranging it effectively to lower the temperature in that state. The molecule's temperature is characterized by the entropy or population dispersion in the ground electronic state; a lowering of the temperature corresponds to a lowering of the entropy by reducing the dispersion amongst the vibrational states. Cooling would ultimately stop when the population had been driven into a single vibrational level.

This ground-state cooling is achieved at the expense of heating a small fraction of the population in an electronic excited state. The coherent radiation of the field provides energy to lower the entropy on the ground-state surface, while raising the corresponding entropy and temperature on the excited-state surface. Except for an initial seed population introduced in the excited state, the suggested cooling

mechanism involves no net transfer out of the ground electronic state.

No claim is made that practical molecular refrigeration will be achieved in this fashion. But this research is a clever example of the latest theoretical schemes for using tailor-made optical pulses to control events at the most intimate atomic and molecular scales.

Control

The dream of molecular control goes back to the early 1960s, when the idea was born of using lasers to manipulate molecular dynamics and/or associated electronic processes, with the goal of selectively breaking chemical bonds (see R. N. Zare, *Nature* **365**, 105; 1993). A long era of interesting but somewhat misguided research followed. Theory largely focused on various issues of molecular dynamics, but never quite came to grips with the problem of designing the control laser. Meanwhile, laboratory researchers ploughed ahead using all manner of existing lasers without proper guidance from theory, and with an inevitable lack of success. Now, with advances on both the theoretical and experimental fronts (for a review, see W. S. Warren, H. Rabitz and M. Dahleh, *Science* **259**, 1581–1589; 1993), the general consensus is that the research is about to bear fruit.

The key theoretical advance made in the past few years is understanding that control of molecular dynamics is inherently a matter of manipulating quantum interference processes. The generic problem may be posed as starting with the molecule in quantum state ψ_i , and steering it to final state ψ_f , in such a way that some molecular objective is met, be it bond-breaking, localized energy deposition or molecular cooling. For experimentalists, the introduction of optimal control techniques at last gives them the tools to address the design of the optical fields. Optimal control theory has its roots in operations research, going back several decades, and the topic has proved to be enormously successful for treating everyday engineering design problems (for example in the flight of aeroplanes). Theorists are now showing that these concepts can just as well be applied to guiding and manipulating the motion of molecules. The past few years have seen frenetic exploration of the concepts of quantum control theory, along with a variety of preliminary illustrations of the control of rotational, vibrational, reactive and electronic degrees of freedom (in some cases simultaneously).

Early approaches largely used continuous-wave or possibly pulsed lasers with

1. Peker, A. & Johnson, W. I. *Appl. Phys. Lett.* **63**, 2342–2344 (1993).
2. Klement, W. *et al. Nature* **187**, 869–870 (1960).
3. Drehman, A. J. & Greer, A. L. *Acta metall.* **32**, 323–332 (1984).
4. Inoue, A. *et al. Mater. Trans. JIM* **31**, 425–428 (1990).
5. Inoue, A. *et al. Mater. Trans. JIM* **33**, 937–945 (1992).
6. Zhang, T. *et al. Mater. Trans. JIM* **32**, 1005–1010 (1991).
7. Saotome, Y. *et al. Proc. 8th Int. Conf. on Rapidly Quenched Metals, Mater. Sci. Engng A* (in the press).

a frequency tuned to some identified fundamental mode or transition. We now know that the problem is far more subtle than this: molecular dynamics will inherently be a complex process, simultaneously involving many frequencies of motion working cooperatively. So the key to experimental progress lies in the rapidly evolving flexible tools of ultrafast laser pulse-shaping. Creating ultrafast pulses and shaping them are two separate arts, and together these tools will meet the needs of molecular dynamical control.

Perhaps most important, theory has identified the need to include laboratory feedback in the optical field design to surmount the problems of uncertainties in realistic molecular hamiltonians and inevitable imperfections in the created control fields. Virtually all problems of practical macroscale engineering control rely on feedback from observation of the current state of the system (for example, an aeroplane's location) to set the controls for the manipulation of the next period of evolution. Although analogous feedback on ultrafast molecular timescales is not possible, an adaptive sequential form of this can be used, as described below.

Pulse-shaping can be computer-controlled, allowing rapid alteration from one pulse to another. Given our inadequate knowledge of molecular hamiltonians for all but the simplest molecules, the designs are unlikely to be spot-on from the start. The answer may be to introduce a sequence of tailored optical pulses and their probes for feedback to a learning algorithm which would rapidly suggest a design for the next experiment, and so forth until the molecules had succeeded in 'teaching' the laser to achieve their control. For this to be practical, stable and reliably shaped pulses are essential (and are becoming available).

My guess is that the next few years will see a basic demonstration of molecular control — the creation of desired final states — by design. Of necessity, initial experiments will be on simple molecules; a likely target is the control of rovibronic states, perhaps including dissociation as well as coherent wavepacket generation. These will serve to confirm the basic principles of molecular control and to test the bounds for manipulation of molecular dynamical processes. Practical applications are expected to follow, and maybe the earlier dream of manipulating chemical reactions will be realized. And, standing the control concept on its head, we might see a new type of molecular spectroscopy designed to make it easy to extract valuable information on intramolecular forces. □

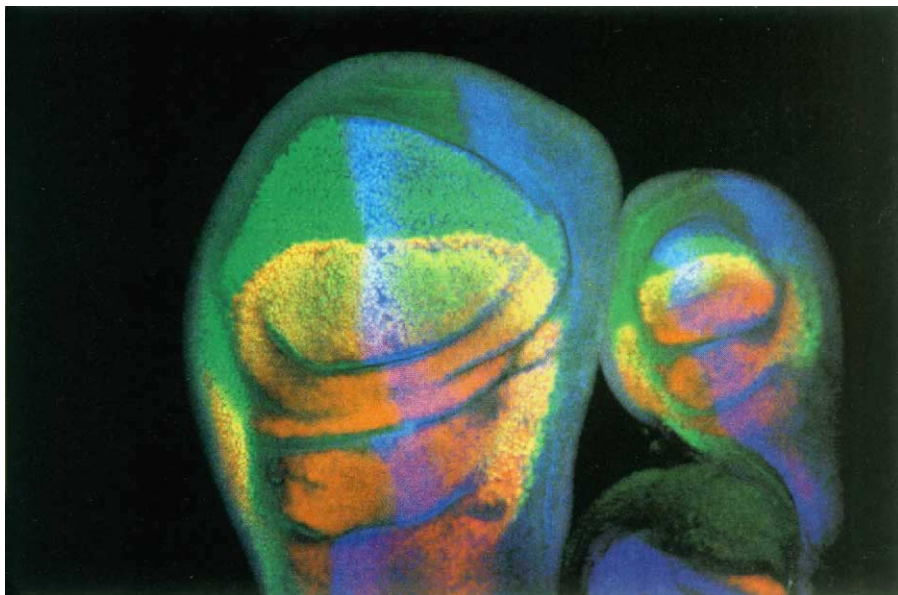
Herschel Rabitz is in the Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.

A no-wing situation

Peter A. Lawrence and Ginés Morata

FEW ideas hold sway for long; without attention their reputations tend to rust. The idea of developmental compartments is an example, but papers by Blair¹ and Diaz-Benjumea and Cohen² have now brought it back to gleaming life. Exactly 20 years ago it was discovered that insects are constructed piecemeal; all the descendants of polyclonal groups of cells that are

into leg 2). But, if some combination of selector genes that is found nowhere in a normal animal, a 'nonsense code word' were created, then the cells would form a messy pattern that would have no normal counterpart⁷. This hypothesis is striking because it does something very few theories do; it provides a link between genetics and anatomy.



Compartments subdivide the wing (left) and haltere (right) imaginal disks of *Drosophila*. One gene (green) is expressed in the appendages (as distinct from the thorax), another (blue) marks the anterior (as distinct from the posterior) compartments, and the other (red) is *apterous*, which is expressed in the dorsal (as distinct from the ventral) compartments. Overlap between red and green gives yellow/orange. (Image courtesy of J. Williams, S. Paddock and S. Carroll.)

picked out in the young embryo generate precisely defined territories (compartments) in the adult. The borders of compartments may coincide with visible landmarks³ or they may strike across featureless terrains⁴.

How could compartments be genetically specified so that each would produce the appropriate piece of the adult? Garcia-Bellido⁵ proposed a brave hypothesis: the compartments would coincide precisely with the realms of action of a special class of 'selector genes', genes that would direct the pattern of development of all the cells within each compartment. Selector genes would be needed continuously from the earliest time when the set of cells in a compartment was founded. The combination of selector genes active in a compartment would constitute a binary code that directs its development⁶: removal of a selector gene from a cell should therefore result in a transformation of that cell and its descendants so that they would make a part of another compartment (for example, changing leg 1

Many parts of the compartment hypothesis were experimentally tested, and, although it is too grand and too simple to be right everywhere and in every detail, it gained much support⁸⁻¹³. But most of the experiments were concerned with the anteroposterior (A/P) axis and with homeotic genes, which partly explains why the wider application of the idea, as imagined in the 1970s, was slow to find favour. For example, Garcia-Bellido and colleagues⁴ had originally proposed that the wing of *Drosophila* is sequentially compartmentalized — the anteroposterior (A/P) subdivision precedes the formation of the imaginal disk but, later on, the anterior and posterior polyclones are both further subdivided into dorsal and ventral (D/V) and proximal (thorax) and distal (wing) polyclones. Indeed it was known that there is a line of lineage restriction that separates the population of dorsal cells from ventral cells in the wing; cells marked in embryos can produce descendants in both dorsal and ventral surfaces, whereas, later, even large clones never

cross over the wing margin and, whether they are dorsal or ventral, meet at a common boundary¹⁴. At the A/P border it is believed that cells do not mix because they have different surface properties; there is thought to be a 'label' on the posterior cells due to the action of the *engrailed* selector gene^{8,9}. But — for both the D/V and thorax/wing compartments — no selector genes were known.

However, in the mid-1980s a zone of cells that divide rather little was discovered near the D/V but not the A/P boundary, and it was suggested that these cells could act as a barrier to the growth of clones across the D/V wing margin¹⁵. If this were the mechanism it would mean that D/V and A/P compartments are made differently. Following this, an influential review¹⁶ discussed the idea that the wing disk is sequentially divided up into true compartments, and then ceremoniously buried it.

Blair¹ and Diaz-Benjumea and Cohen² have opened up this particular grave. They show, with beautiful evidence, that the wing disk is divided into dorsal and ventral compartments by a selector gene. The gene is *apterous*, which encodes a transcription factor containing both a homeodomain and a LIM domain¹⁷. In *apterous*⁻ flies, wings and halteres (homologous to the hind wings of other insects) fail to develop, but the thoracic trunk is almost normal; the remaining parts of the fly are unaffected. The *apterous* protein is found in the dorsal region of the wing disk from early second instar when the disk contains fewer than 200 cells¹⁸. This is at the time when the D/V boundary comes into existence, as defined by cell lineage¹². Both the borders of marked clones and the limit of *apterous* expression coincide exactly for many cells.

Blair demonstrates that this boundary runs along the middle of the zone of 'non-proliferating' cells previously proposed to act as a barrier; the dorsal clones (which express *apterous*) include one half of the zone, and the ventral the other — results that comprehensively disprove the barrier hypothesis. Furthermore, the dorsal expression of *apterous* is found in both anterior and posterior compartments, even though these polyclones were separated in the young embryo and belong to different parasegments. This illustrates the way in which selector genes combine to specify development of compartments.

Diaz-Benjumea and Cohen also made clones of cells that lack the *apterous* gene. In the ventral compartment, the clones develop normally — hardly surprising, because the gene is not expressed there. But in the dorsal compartment, the *apterous*⁻ clones make apparently perfect ventral cells. If they are initiated when the disk contains only tens of cells, they usually join their ventral colleagues and become subsumed into the ventral com-

partment when it is formed. If they are made later and away from the D/V border, they grow as patches of ventral cells in a dorsal sea; new D/V boundaries, with characteristic bristles, form around their perimeters.

The creation of a new and circular border has a fascinating consequence: it induces a localized outgrowth of the wing with a normal proximodistal pattern. An organizing role for the border in growth had been noted before⁹ but, as Diaz-Benjumea and Cohen explain, this new observation helps explain why the wing itself grows out. It seems likely that, in normal development, the creation of dorsal cells, which is associated with the activation of *apterous* in the second larval stage, is in itself enough to trigger the formation of a compartment border, with all its special properties. The border probably represents the upper (or lower) limit of a gradient field. If such a limit (say 10) is formed in a part of the field at a different value (say 7) there will be a blending interaction between border and surround, leading to closely packed intermediate values (8, 9) and therefore rapid growth¹⁹.

Finally, the two new papers should invigorate the search for compartments elsewhere in the fly and, more importantly, in other animals. In our opinion the compartment hypothesis has not been sufficiently applied outside insects, for example to the development of vertebrates. Even so, the rhombomeres in the vertebrate hind-brain are being imaginatively investigated and the evidence that they are compartments — both as units of cell lineage²⁰ and selector gene action²¹ — is accumulating nicely. □

Peter A. Lawrence is in the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. Ginés Morata is at the Centro de Biología Molecular, Universidad Autónoma de Madrid, Madrid 28049, Spain.

1. Blair, S. S. *Development* **119**, 339–351 (1993).
2. Diaz-Benjumea, F. & Cohen, S. *Cell* **75**, 741–752 (1993).
3. Lawrence, P. A. *Nature New Biol.* **242**, 251–253 (1973).
4. Garcia-Bellido, A., Ripoll, P. & Morata, G. *Nature New Biol.* **245**, 251–253 (1973).
5. Garcia-Bellido, A. *CIBA Fdn. Symp.* **29**, 161–182 (1975).
6. Garcia-Bellido, A., Lawrence, P. A. & Morata, G. *Scient. Am.* **241**, 102–110 (1979).
7. Struhl, G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7380 (1982).
8. Morata, G. & Lawrence, P. A. *Nature* **255**, 614–617 (1975).
9. Lawrence, P. A. & Morata, G. *Dev. Biol.* **60**, 321–337 (1976).
10. Steiner, E. *Roux Arch. dev. Biol.* **180**, 9–30 (1976).
11. Struhl, G. *Nature* **270**, 723–725 (1977).
12. Morata, G. & Lawrence, P. A. *Dev. Biol.* **70**, 355 (1979).
13. Kornberg, T. *Dev. Biol.* **86**, 363–372 (1981).
14. Bryant, P. J. *Dev. Biol.* **22**, 389–411 (1970).
15. O'Brochta, D. A. & Bryant, P. J. *Nature* **313**, 138 (1985).
16. Brower, D. *Cell* **41**, 361–364 (1985).
17. Cohen, B., McGuffin, M. E., Pfeiffe, C., Segal, D. & Cohen, S. M. *Genes Dev.* **6**, 715–729 (1992).
18. Williams, J. A., Paddock, S. W. & Carroll, S. B. *Development* **117**, 571–584 (1993).
19. Lawrence, P. A. *The Making of a Fly* (Blackwell Scientific, Oxford, 1992).
20. Guthrie, S. *et al.* *Nature* **356**, 157–159 (1992).
21. Guthrie, S., Prince, V. & Lumsden, A. *Development* **118**, 527–538 (1993).

Solar wind

A NEW defence against Earth-impacting asteroids has recently been proposed by H. J. Melosh and I. V. Nemchinov (*Nature* **366**, 21; 1993). A huge solar mirror, flying alongside the asteroid, could focus sunlight to a point on its surface. The resulting steady plume of evaporation would slowly perturb the asteroid's orbit away from fatal impact with the Earth.

Daedalus is taking this idea a step further. Suppose, he says, that the exact sub-solar point of the asteroid could be suddenly heated into such a hotspot. The resulting plume of vapour would form a refractive volume in space, a sort of gas-lens with a cylindrical axis of symmetry pointing towards the Sun. It would tend to bend sunlight inwards towards the hotspot, mirage-fashion. Lenses like this, which focus by means of refractive-index gradients, are well known to biologists. Our own eyes use this strategy; and certain fishes have eyes whose rod-like lenses focus almost entirely by their cylindrical refractive symmetry. If the plume of expanding gas from the asteroidal hotspot happened to focus the sunlight exactly onto its source, it would be self-maintaining.

DREADCO's physicists are working out the details. What temperature, shape and depth makes a hotspot self-maintaining? How fast can it move over the surface of a spinning asteroid, always tracking the sub-solar point? If a small perturbation defocuses the sunlight, under what conditions will the plume react to refocus it? From the answers, Daedalus hopes to devise the ultimate anti-asteroid strategy. The celestial menace will be attacked, not with a nuclear bomb or a vast solar mirror, but with a precise charge of smokeless powder fired exactly at its sub-solar point. The resulting plume of expanding gas will focus sunlight onto its source, and keep the evaporation going. Under the steady thrust of its new solar-powered rocket motor, the asteroid will slowly diverge away into a non-threatening orbit.

This elegant scheme also lends itself to propelling spacecraft, at least away from the Sun. A small craft made of a volatile material like ammonium chloride might capture enough solar energy to accelerate quite rapidly. But as a terrestrial defence, the idea should be tried out first on a harmless asteroid. As the white-hot crater or equatorial channel of rotation dug its way into the object, its changing brightness and emission spectrum could identify the material being evaporated and monitor its motor action. It would be rank bad luck if the propulsive effect of the hotspot set the asteroid on a collision course with the Earth.

David Jones

Extinction or 'co-extinction' rates?

SIR — Defining precisely when a species of bird or mammal has become extinct is notoriously difficult¹. The task is even more challenging for smaller organisms. To date only 61 species of insect are known to have become extinct since 1600 (ref. 2), yet there are about 10^6 – 10^7 species in this group^{3–5}. Smith *et al.* attempted to overcome the problem of measuring extinctions by using changes in the threatened status of species to provide new estimates of extinction rates⁶. The

fleas sometimes less so⁸. In both cases there are many insects restricted to single species of host, and as a host species becomes extinct, so does one or more species of parasite. Many will have shed a tear for the passing of the last passenger pigeon in 1914 (ref. 2) but until now the similar demise of at least two species of passenger pigeon louse (*Columbicola extinctus* Malcomson and *Campanulotes defectus* Tendeiro) has been ignored (see figure). How many species of lice and fleas

sider the value of insects, either positive or negative (with perhaps the exception of butterflies), solely in terms of their contribution to ecosystems and their direct or indirect effect on humans. When discussing conservation priorities and practices, we need to examine these values rather than simply condemning portions of the animal kingdom to extinction without thought. There may be conflicts in conservation needs, forcing us to bid farewell to the gorilla louse or the lice of the Californian condor while retaining their hosts. If so, we should do so in the full knowledge of what is being lost.

Nigel E. Stork

Christopher H. C. Lyal

Department of Entomology,
The Natural History Museum,
Cromwell Road, London SW7 5BD, UK



Left, the passenger pigeon *Ectopistes migratorius* (Linnaeus), from ref. 9 (NHM picture library). Above, passenger pigeon louse *C. extinctus* Malcomson (P. York/NHM).

measures they used suggest that the characteristic extinction time for half the species of birds and mammals is 200–300 years, and that for invertebrates such as molluscs is 60,000 years. There were insufficient data to provide a similar estimate for insects.

We believe that at least one group of insects, lice (Phthiraptera), and possibly another, fleas (Siphonaptera), may have a similar extinction rate to mammals and birds. Lice are extremely host-specific⁷,

have become extinct since 1600? In addition, how many internal parasites, both multicellular and protist, have vanished with their hosts?

This process of 'co-extinction' is problematic. While zoos concentrate increasingly on the conservation of threatened mammals and birds, the plight of the lice and fleas on these animals is ignored. Indeed, attempts are often made to sanitize the threatened vertebrates, and who can say whether this process causes more insect extinctions? Perhaps more fundamentally, these examples of co-extinction result from close co-evolution⁷, which itself can be viewed as one extreme of a continuum of species associations. In any ecosystem, extinction of one keystone species may lead to the loss of many others. Inevitably, lists of species extinctions and threatened species fail to take this into account.

Those who would not consider a louse to be of the same value as a bird raise an important ethical issue. Most people con-

Hormone signal response system

SIR — A property of signal-response systems in general, and of G-protein-coupled receptors in particular, is that prolonged stimulation often results in reduced responsiveness to further challenge by the same stimulus. Receptor desensitization in humans is responsible for habituation to light, odours, chemical stimulants and narcotics, and is a factor that limits the efficacy of many therapeutic agents. As described for the β -adrenergic receptors, desensitization most certainly includes rapid receptor phosphorylation at cytoplasmic domains of the protein (for example by cyclic AMP-dependent protein kinase, PKA) and uncoupling from the G protein^{1,2}.

Namba *et al.*³ describe four alternatively spliced receptor isoforms of the prostaglandin E receptor, each having distinct second-messenger signalling properties. Data presented in their paper suggest another important functional difference between two of the receptor variants with regard to desensitization. Both the EP3B and EP3C isoforms were shown to mediate cAMP accumulation in transfected CHO cells, although EP3B had a significantly lower EC₅₀ for the agonist. In cells pretreated with forskolin (which elevates cAMP production), however, this order of activity is reversed. The molecular basis for this change could result from differences in receptor desensitization. EP3B has nine Ser/Thr residues in the C-terminal 'tail' domain, at least one of which represents a potential PKA recognition site⁴, and may therefore be

1. Diamond, J. *Conserv. Biol.* **1**, 77–79 (1987).
2. World Conservation Monitoring Centre *Global Biodiversity: Status of the Earth's Living Resources* (Chapman & Hall, London, 1992).
3. Stork, N. E. *Biodiv. Conserv.* **2**, 215–232 (1993).
4. Hammond, P. M. in *Global Biodiversity: Status of the Earth's Living Resources* (Chapman & Hall, London, 1992).
5. May, R. M. *Scient. Am.* **268**, 146–149 (1993).
6. Smith, F. D. M., May, R. M., Pellew, R., Johnson, T. H. & Walter, K. S. *Nature* **364**, 494–496 (1993).
7. Lyal, C. H. C. in *Coevolution and Systematics* (Clarendon, Oxford, 1986).
8. Marshall, A. G. *The Ecology of Ectoparasitic Insects* (Academic, London, 1981).
9. Audubon, J. J. L. *Birds of America* (Macmillan, New York, 1839).

1. Dohlgan, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. *A. Rev. Biochem.* **60**, 653–688 (1991).
2. Hausdorff, W. P., Caron, M. G. & Lefkowitz, R. J. *FASEB J.* **4**, 2881–2889 (1991).
3. Namba, T. *et al. Nature* **365**, 166–170 (1993).
4. Pearson, R. B. & Kemp, B. E. *Meth. Enzym.* **200**, 62–81 (1991).

phosphorylated after forskolin treatment. EP3C has only one Ser and does not conform to the PKA consensus sequence. Thus the work by Namba *et al.*³ may provide insights into the mechanisms of negative, as well as positive, regulation of G-protein-coupled receptors.

Henrik G. Dohlman

Department of Pharmacology,
Yale University School of Medicine,
PO Box 9812,
New Haven,
Connecticut 06536, USA

Water in life

SIR — Segel and Tyson in Correspondence¹ tried to defend Maddox's assertion² that molecular biology lacks a quantitative dimension. Maddox then went on to discuss³ an example of an attempt to redress the balance in describing a quantification of the hydration of glucose⁴.

In describing the interaction of glucose and water, the assumption in classical biochemistry has been that metabolically active glucose is free within the pathways of the cell and that water is just 'ordinary' free water⁵⁻⁷. When a calculable theory is available, quantitative considerations fall easily into place, that of classical biochemistry being the law of mass action.

Maddox's suggestion in the first of his two articles mentioned above² was to resurrect the law of mass action, but the past 20 years have shown that this may apply only following cataclysmic cell disruption to yield aqueous solutions *in vitro*⁸, and is generally inappropriate to the interior of the cell *in vivo*. It is because no adequate or simple alternative has been found that molecular biology languishes in its role as a descriptive discipline.

Maddox's second article on this topic³ finished with the words "real water". The problem of describing the physical state of water is extremely complex⁷, and as yet no calculable theory exists with which to begin to consider water's interactions with other molecules. Albert Szent-Györgyi himself said⁹ that "life is water dancing to the tune of the solids".

Denys N. Wheatley

Department of Pathology,
University of Aberdeen,
Aberdeen AB9 2ZD, UK

1. Segel, I. & Tyson, J. J. *Nature* **357**, 106 (1992).
2. Maddox, J. *Nature* **355**, 201 (1992).
3. Maddox, J. *Nature* **364**, 669 (1993).
4. Cramer, C. J. & Truhlar, D. G. *J. Am. chem. Soc.* **115**, 5745-5753 (1993).
5. Nengendank, W. & Edelman, L. (eds) *The State of Intracellular Water* **2**, No. 2, 1-114 (1988).
6. Clegg, J. S. *Collective Phenomena* **3**, 289-312 (1981).
7. Ling, G. N. *A Revolution in the Physiology of the Living Cell* (Krieger, Malabar, Florida, 1992).
8. McConkey, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3236-3240 (1982).
9. Szent-Györgyi, A. *Perspect. Biol. Med.* **14**, 239-249 (1971).

Death and c-fos

SIR — Smeyne *et al.*¹ suggest that continuous expression of c-fos precedes programmed cell death (apoptosis) *in vivo*. They do not show, however, that *fos-lacZ* is expressed in the specific cells that go on to die *in vivo* and the developmental examples they choose do not exclude other explanations for c-fos expression.

First, their tooth example (Fig. 1d of ref. 1) discusses and illustrates the so-called enamel knot. But the structure expressing the *lacZ* marker they show is not the enamel knot but the dental papilla mesenchyme. The enamel knot is a transitory structure present at the late cap stage of development within the enamel organ². The *fos-lacZ* labelled cells of the dental papilla mesenchyme do not normally undergo programmed cell death, but differentiate into the cells that will form dentine (odontoblasts) and the adult dental pulp.

Their second example is the midline epithelium of the secondary palate¹. Medial edge epithelial cells of the palate, however, do not die, but rather migrate³; indeed, the reference quoted by Smeyne *et al.*¹ makes this very point⁴! Further, the c-fos expression patterns in the palate are not consistent with c-fos expression preceding programmed cell death. For many years, it has been known that the biochemical changes which precede medial edge epithelial cell differentiation (death?) occur about 24 hours before the event (see ref. 5 for review). However, Smeyne *et al.*¹ illustrate c-fos expression only where the medial edge epithelia touch and fuse to form a midline epithelial seam (their Fig. 1e). Where the medial edge epithelia are not in contact (Fig. 1f) there is no c-fos expression. However, these events are separated by only a few hours. If c-fos is a harbinger of death, one would expect to see the expression about 24 hours previously.

Interestingly, these patterns of *fos-lacZ* expression in the developing tooth and palate¹ resemble those of various cytokines, particularly transforming growth factor- β family members^{6,7}. In the palate, transient expression of transforming growth factor β 3 in the medial edge epithelia⁶ matches closely that reported by Smeyne *et al.* for *fos-lacZ*. Prolonged c-fos expression in development may therefore correlate better with cytokine signalling at sites of cellular interactions

than with imminent programmed cell death.

Mark W. J. Ferguson

School of Biological Sciences,
University of Manchester,
Manchester M13 9PT, UK

MORGAN AND CURRAN REPLY — Ferguson suggests that we do not show that cells expressing *fos-lacZ* go on to die *in vivo*. But Fig. 2 of our paper shows *fos-lacZ* staining in the cytoplasm and remnants of hippocampal neurons following treatment with the neurotoxin kainic acid. We cited several examples of the association between *fos-lacZ* expression and cell death *in vivo*, but not that the *fos-lacZ*-positive cells went on to die in all cases. We had to remove an additional figure demonstrating further examples of the association because of *Nature's* space constraints. We did not wish to imply that *fos-lacZ* expression occurs exclusively in dying cells; indeed, we have previously documented c-fos expression in proliferating and differentiating cell populations.

Ferguson further suggests that the developmental examples we chose are problematic. Unfortunately, for reasons of space we had to delete data on *fos-lacZ* expression in the periderm, interdigital web cells and hypertrophic chondrocytes; all circumstances in which developmentally programmed cell death is documented. Expression of *fos-lacZ* in the medial edge epithelium of the palate is indeed a complex phenomenon, and our discussion of this point was also curtailed for the sake of brevity. It is certainly of great interest to determine the fate of *fos-lacZ*-positive cells in this structure.

Regarding *fos-lacZ* expression in the developing tooth, we have published a correction¹ about our misidentification of the enamel knot. We thank A. Lumsden and I. Thesleff for bringing this to our attention, and apologise again for any confusion this may have caused.

The association of *fos-lacZ* expression with TGF β family members was also omitted from our paper for space reasons, and is mentioned in ref. 8. Ferguson may wish to consider the possibility we are now investigating, that cytokines such as TGF β may actually contribute to cell death.

J. M. Morgan

J. Curran

Roche Institute of Molecular Biology,
Nutley,
New Jersey 07110, USA

1. Smeyne, R. J. *et al.* *Nature* **363**, 166-169 (1993); correction **365**, 279 (1993).
2. MacKenzie, A., Ferguson, M. W. J. & Sharpe, P. T. *Development* **115**, 403-420 (1992).
3. Carette, M. J. M. & Ferguson, M. W. J. *Development* **114**, 379-388 (1992).
4. Shuler, C. F. *et al.* *Dev. Biol.* **154**, 318-330 (1992).
5. Ferguson, M. W. J. *Development* **103**, Suppl. 41-60 (1988).
6. Fitzpatrick, J. *et al.* *Development* **109**, 585-595 (1990).
7. Varahokari, A. *et al.* *Development* **113**, 985-994 (1991).
8. Molinar-Rode, R. *et al.* *Molec. cell. Biol.* **13**, 3213-3220 (1993).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

The cosmic distance scale and the Hubble constant

M. Fukugita, C. J. Hogan & P. J. E. Peebles

Although astronomers can reliably measure the relative distances of distant galaxies, attempts to calibrate the absolute cosmic distance scale remain controversial. Nevertheless, the sources of possible error are now clearly defined and a convincing result seems to be within reach.

IN a perfectly uniform and isotropic expanding universe, particles move according to the Hubble law, $v = H_0 r$, where v and r are the relative velocity and separation of any two particles. The value of the Hubble constant, H_0 , fixes the fundamental scale of length and time of the cosmic expansion. Precise knowledge of H_0 is needed for a quantitative test of virtually any aspect of cosmological theory: it bears on the predicted age of the Universe, the problem of the cosmological constant, comparisons of Big Bang nucleosynthesis with observations of matter density, predictions for dark-matter mass density and the confrontation of structure formation theories with anisotropies in the thermal cosmic background radiation.

The global value of H_0 has long been uncertain by about a factor of two^{1,2}, largely because absolute distances are easily measured only on small scales, where inhomogeneous gravitational acceleration causes motions to depart significantly from the simple expansion described by the Hubble law. The uncertainty arises from the disagreement of different techniques used

to connect the local distances to the smooth large-scale Hubble flow.

On local scales, there is a consensus regarding the distance to the nearest galaxy outside our own, the Large Magellanic Cloud (LMC), at $r = 50 \pm 3$ kpc. At least four independent measures agree; the expansion of the photosphere of supernova 1987A (ref. 3), the parallax of its light echo⁴, and the two types of well characterized variable stars whose absolute luminosity can be inferred from their period, Cepheids and RR Lyrae stars⁵. The LMC is so close that it simply orbits our Galaxy; its radial velocity has nothing to do with the Hubble expansion. Nevertheless, the LMC measurements provide a validation of the Cepheids, which then give reliable distances to other nearby 'calibrator' galaxies. The most distant reliable calibrator is M81, with a distance (now accurately measured by the Hubble Space Telescope⁶) of 3.63 ± 0.34 Mpc.

The convergence to uniform Hubble flow on the very largest scales is most clearly demonstrated using the (remarkably

FIG. 1 Images of the small local galaxy IC4182. The two upper frames show the galaxy. The box in the left frame is enlarged to give the right frame; the box in the right frame shows the area enlarged in the two lower frames. The arrows in the lower frames show a Cepheid variable star in this galaxy at a bright and a faint stage in its luminosity variation. For such stars, the period of the light variation gives an accurate estimate of its intrinsic brightness and therefore of its distance. This galaxy is of special importance because it was the site of a type Ia supernova, 1937C. Measurements last year of 27 Cepheid variables in this galaxy by the Hubble Space Telescope¹³ (HST) gave the first reliable distance estimate for IC4182, and hence the first reliable direct determination of intrinsic brightness for a type Ia supernova. This calibration gives the distance to other supernovae farther away, allowing a direct estimate of the Hubble constant. The HST has now measured Cepheids in other galaxies, including another supernova host galaxy (NGC5253) and the important local calibrator M81. A key project is underway to measure direct Cepheid distances to galaxies in the Virgo galaxy cluster, eliminating some of the most controversial intermediate stages in establishing the cosmic distance scale. (Images courtesy of Hubble Space Telescope News, Space Telescope Science Institute.)

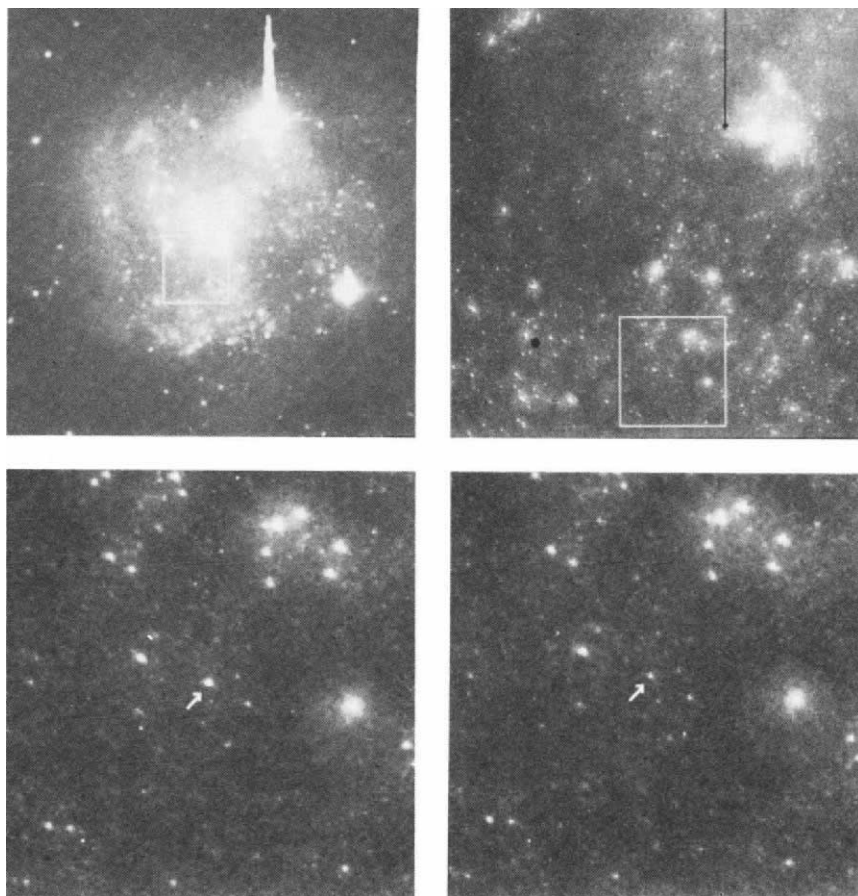
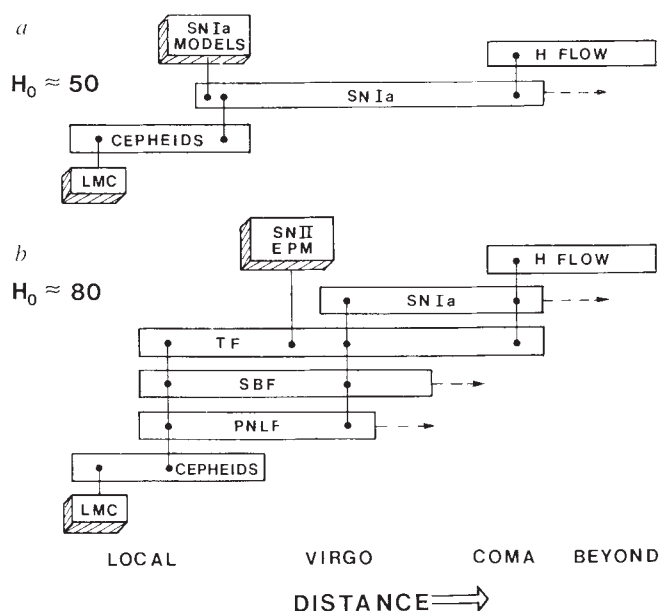


FIG. 2 *a, b*, Two cosmological distance ladders. Horizontal bars indicate the distance range for each method as discussed in the text, vertical bars indicate direct links between methods. Absolute calibration is established through Cepheid variable-star distances, by models of the type Ia supernova luminosity (SN Ia), or model fits to the expanding photospheres of type II supernovae (SN II). Cepheid calibrations are currently available only for galaxies in or near the Local Group of galaxies. Precise measurement of H_0 however requires connecting these to the large-scale smooth Hubble flow. (LMC, Large Magellanic Cloud; TF, Tully–Fisher; SBF, surface brightness fluctuations; PNLF, planetary nebula in a galaxy; EPM, expanding photosphere method; H FLOW, Hubble flow.)



uniform) brightest galaxies in clusters^{7,8}, which show that fractional departures from the Hubble law $\delta v/v \lesssim 10\%$ for $v \gtrsim 10,000 \text{ km s}^{-1}$. Distance ratios between clusters obtained from the Tully–Fisher relation⁹ (see below) and other techniques¹⁰ tie individual galaxy clusters to this cosmic reference frame and reliably establish their true ‘Hubble velocity’, separated from peculiar non-Hubble motions. For example, the Hubble velocity at the distance of the Coma cluster of galaxies is determined to be $v \approx 7,100 \text{ km s}^{-1}$. If we only knew the distance to the Coma cluster, the value of H_0 would be accurately determined.

The nearest cluster tied into this system is the Virgo cluster of galaxies, whose distance ratio to Coma (~ 0.18) is known to $\sim 10\%$ accuracy. Again, if we knew the absolute distance to Virgo, we would know the global value of H_0 without problems connected with peculiar velocities.

But the distance to Virgo is uncertain by a large factor: different arguments give a distance to the cluster core of between 14 and 22 Mpc. The current debate, and our discussion, thus focus on the absolute calibration of a few of the most reliable distance measures which bridge the gap at intermediate distances between a few megaparsecs and a few tens of megaparsecs—that is, between the local reliable calibrations of distance (such as M81) and the distant reliable calibrators of Hubble velocity (such as the Virgo cluster). Although this gap is closing, current techniques appear to give dichotomous results, even allowing for their internal error estimates. We will discuss in turn the case for a ‘low’ H_0 , of $50 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and a ‘high’ H_0 , of $80 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$, with the critical links for each shown schematically in Fig. 2.

The longer distance scale: $H_0 = 50 \pm 10$

The most reliable link leading to a low Hubble constant is provided by the supernovae classified as type Ia (on the basis of their characteristic spectra and, more often, the time history of the luminosity). The Hubble diagram of recession velocity against relative distance inferred from observed maximum supernova brightness has a fairly narrow dispersion at large distances, showing that the standard deviation in supernova luminosity is $< 40\%$, and possibly better with suitable selection^{11,12}; these supernovae are therefore reasonably good ‘standard candles’.

In an important recent development¹³, the Hubble Space Telescope measured Cepheid variable stars in the host galaxy, IC4182, of the third supernova discovered in 1937; SN1937C, a prototypical type I supernova (Fig. 1). If SN1937C has a stan-

dard type Ia luminosity, the distance to IC4182 (4.8 Mpc) fixes the distances of remote type Ia supernovae, giving a low value of H_0 . Moreover, the intrinsic luminosity of SN1937C implied by this distance is consistent with theoretical predictions from the standard model for type Ia supernovae^{14,15}. (In this model a white dwarf star made of carbon or oxygen explosively burns to the iron group when it reaches the Chandrasekhar mass of ~ 1.4 solar masses by accretion from a nearby star. Once one accepts this idea, the maximum total luminosity follows from the mass of radioactive nickel produced.) Thus two ways to establish the absolute luminosity of type Ia supernovae give the same answer.

It follows that if the longer distance scale were wrong, then two independent ways to calibrate type Ia supernovae (Fig. 2a) would have to be wrong yet fortuitously agree. For example, if $H_0 = 80 \text{ km s}^{-1} \text{ Mpc}^{-1}$, then SN1937C would have to be the most (intrinsically) luminous type Ia supernova on record—about 2.5 times as bright as average, or almost 3 standard deviations. This is not to be excluded: for example, if SN1937C, instead of being typical, were as intrinsically luminous as the brightest of the dozen or so well studied distant type Ia supernovae, we would get $H_0 = 69 \pm 8 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Second, if $H_0 = 80 \text{ km s}^{-1} \text{ Mpc}^{-1}$, then the typical type Ia supernova would have to be less than half as bright as predicted by the standard Ia supernova model (the agreement with SN1937C would still hold, as the distance to IC4182 is fixed). This is also conceivable; a number of recent modifications^{16,17} to the standard model, including improved models of the opacity, radiative transfer, and the explosive deflagration/detonation process, partly motivated by observed light curves, ejecta velocity and nucleosynthesis products, tend towards a higher value of H_0 .

The shorter distance scale: $H_0 = 80 \pm 10$

A sensible case can also be made for a high value of H_0 ^{18,19}. As indicated in Fig. 2b, there are at least four independent methods for measuring relative distances to galaxies, and two independent empirical methods of attaching an absolute distance to these distance ratios which consistently give $H_0 = 80 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

One way to establish distance ratios uses the Tully–Fisher (TF) relation between the luminosity of a spiral galaxy and its rotation speed²⁰; more rapidly rotating galaxies tend to be more massive, and hence more luminous. Like its somewhat less precise sister relation for elliptical galaxies (the Faber–Jackson or

D_n - σ relation relating size D_n and velocity dispersion σ), TF reflects poorly understood but apparently precise regularities in the ways galaxies happen to have formed: over a wide range of masses, galaxies exhibit narrowly defined correlations between mass-to-light ratio and surface brightness. Because rotation velocities are determined from the Doppler shift in galaxy spectra (usually 21-cm linewidth), which can be measured at large distances, the TF relation can be used to predict relative galaxy luminosities and hence distance ratios. The Tully-Fisher relation gives a reliable distance ratio between two galaxies with an empirical accuracy of $\sim 15\%$ per galaxy¹⁸, and establishes even more accurate distance ratios between galaxy clusters⁹. A large dispersion in the TF distances to spiral galaxies in the direction of the Virgo cluster, which had cast doubt on the viability of this technique, is now known to arise from the true depth of this cluster in space^{18,21}. The true scatter of TF is remarkably narrow, so that the notorious 'Malmquist bias' which formerly plagued this method is small, and is now statistically controlled²². (Malmquist bias occurs if a sample systematically favours intrinsically brighter galaxies at greater distance.)

A check on the reliability of TF comes from two relatively new techniques, which have matured in the past few years to equal or even surpass the precision of TF. One technique uses surface brightness fluctuations (SBF)²³ in galaxies, in which the statistical variance in surface brightness is adopted as a measure of the distribution of star luminosities in a galaxy. This can be used at large distances, even when the individual stars cannot be spatially resolved because of atmospheric and optical blurring. The method has been applied to several hundred galaxies, and, with corrections for variations in stellar populations, yields relative distances accurate to $\sim 5\%$ per galaxy. Figure 3a shows the remarkable consistency of distance ratios for galaxies where both the TF and SBF methods have been applied.

The second new method uses the shape of the luminosity function for the planetary nebulae in a galaxy (PNLF)²⁴, which is found to display a sharp upper cutoff in luminosity. Observational evidence²⁴ shows that the cutoff is surprisingly universal. Again, PNLF distance ratios are in excellent agreement with SBF and TF, and shown in Fig. 3b. For all these techniques, TF, SBF and PNLF, the empirical precision is remarkably good, even better than one might have anticipated from first principles. Their precision is even confirmed by comparison with distance ratios from type Ia supernovae, although they clearly disagree with the absolute brightness calibration using the methods described above.

The value $H_0 = 80 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$ then follows if one uses Cepheid distances to a half-dozen nearby calibrator galaxies²⁵ for which one or more of these methods can be applied. Alternatively, instead of relying on a few local galaxies calibrated by Cepheids, we can anchor TF distances with a direct physical measure of distance called the expanding photosphere method (EPM)³. This method uses type II supernovae, produced by core collapse during the death throes of massive stars (SN1987A being the best known example). These are not standard candles like type Ia supernovae, but their distances can be estimated using models that fit measured spectral properties of their ejecta. Like the older Baade-Wesselink method applied to variable stars, EPM yields an absolute distance by comparing the physical velocity of the ejecta (measured from the Doppler shift of absorption lines in the gas) to rate of change in angular size of the material (derived from the temperature and brightness, and measured at several epochs). The EPM requires a model atmosphere to enable the calculation of the emergent flux, for the photosphere does not radiate like an ideal black body. It also requires a tricky estimate of temperature from observed colours. Although the correction is substantial and the colour-temperature identification contains some uncertainties, the calculation yields an accurate distance to SN1987A in the LMC³, and the stability of the estimated distances as the supernovae expand and cool is impressively good. In at least seven of the ten most carefully studied type II supernovae, EPM distances agree well with TF distances to the host galaxy calibrated with Cepheids²⁶, independently supporting a high value of H_0 . (TF is needed here because the host galaxies of these type II supernovae happen to lie in a local region of slower than average expansion, so that an application of EPM on its own underestimates H_0 .)

Thus, if $H_0 = 50 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$, all of the following are required: the TF distance for the six nearby calibrators needs to be anomalous, PN and SBF techniques need to be anomalous for two of these (M31, M81) and the EPM technique must be systematically wrong in the seven best studied cases, except in the independently validated cases, SN1987A and SN1970g, whose EPM distances are confirmed by Cepheid distances³. Any of these is a possibility, but the required anomalies are much larger than expected from empirically measured dispersions.

The age of the Universe

The value of H_0 has an important effect on cosmological theory, in particular on the relativistic, homogeneous expanding universe models which relate expansion rate to cosmic age. For

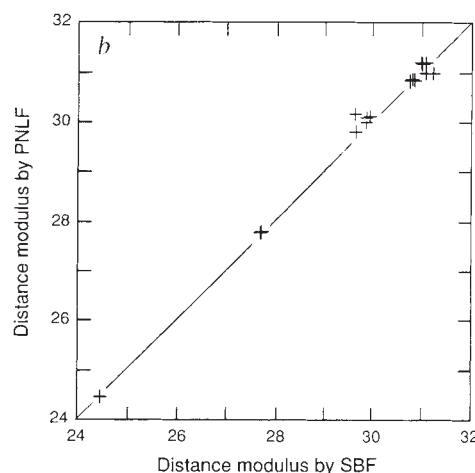
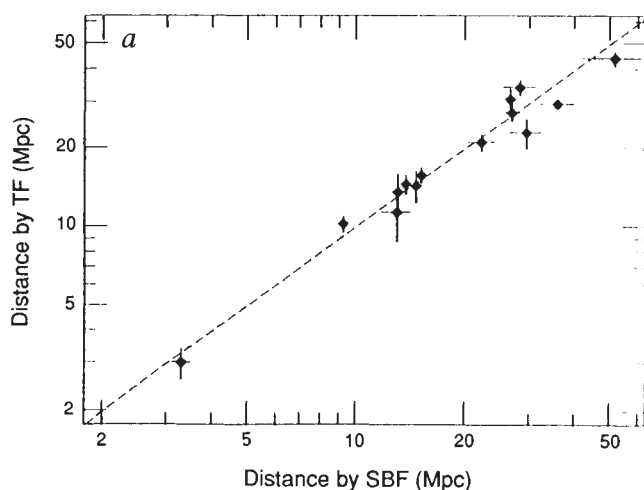


FIG. 3 a, Comparison of galaxy distances determined by the Tully-Fisher (TF) and surface brightness fluctuation (SBF) methods. Part of Fig. 29 from Jacoby *et al.*¹⁸ by permission. b, Comparison of galaxy distances determined by the surface brightness fluctuation (SBF) and

planetary nebula luminosity functions (PNLF) methods, for 16 galaxies. The logarithmic scale here is the distance modulus $5 \log (d/10 \text{ pc})$; for instance 25 stands for 1 Mpc, 32 for 25 Mpc. (Data from G. Jacoby, R. Ciardullo and J. Tonry, personal communication).

$H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ the Hubble time, given by H_0^{-1} , is 20 Gyr. This is the age of the expanding universe if the expansion 'coasts', as in a cosmological model in the limit of very low mass density. The real Universe of course contains matter, which tends to decelerate the expansion by its gravity, and may also contain a gravitating physical vacuum (or 'cosmological constant' Λ), which accelerates the expansion. On grounds of elegance, most physicists would prefer the Einstein-de Sitter model, in which $\Lambda = 0$ and the mass density is such that the effective gravitational potential energy per unit mass is equal to the kinetic energy. (This means that the Universe expands at precisely its escape velocity, and, in relativistic cosmology, that space curvature is negligibly small.) In the Einstein-de Sitter model, the age of the Universe from high density to now is two-thirds of the Hubble time, or about 13 Gyr if $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$. This is comparable to the ages of the oldest stars, as derived from models for stellar evolution. On the other hand, the higher value $H_0 = 80 \text{ km s}^{-1} \text{ Mpc}^{-1}$ would make the Hubble time $H_0^{-1} = 12 \text{ Gyr}$ and the Einstein-de Sitter age only 8 Gyr, too small by any accepted measure: one would have to learn to live in a low-density Universe, with an age of $\sim 11 \text{ Gyr}$ if $\Lambda = 0$ and the mass density is 10% of the Einstein-de Sitter value (the minimum implied by dynamical estimates³⁶). This would still not be welcomed by those who study stellar evolution, who consider it difficult to push the maximum ages below 13 Gyr. It would be abhorred by those who study the inflation picture as a possible explanation for the large-scale isotropy of the visible Universe, for that requires space curvature be negligibly small, as in the Einstein-de Sitter model.

A low-density Universe with a non-zero cosmological constant can have negligible space curvature (as demanded by inflation) and an expansion timescale longer than the Hubble time; however, in such models the predicted frequency of gravitational lensing of distant quasars by galaxies positioned close to the line of sight is considerably larger than is observed²⁷. One way out of this constraint is that at the large distances typical for lensing events, galaxies are often observed to have considerably larger star formation rates than in their nearby older counterparts, and might also contain larger quantities of dust; lensing events could therefore be obscured. If so, the reduced rate of observed lensing events might be consistent with the low-density cosmologically flat Universe which is the second choice for those who study inflation, and which may thus offer the easiest way to reconcile the shorter distance scale with the ages of stars and the elements. Confirmation of non-zero cosmological constant—a gravitating physical vacuum—would have important ramifications for fundamental physics.

An end to the controversy?

We have focused so far on the currently most precise and best

statistically controlled methods. There are, however, other methods of fixing the absolute distance scale^{18,19,28}, of which we list a few particularly active cases. Recently, high-resolution ground-based imaging has revealed variable stars in some Virgo cluster galaxies²⁹. Some of these are long-period variables, and some may be Cepheids; the current data are still too sparse to be certain, but the fact that they are bright enough to be seen at all tends to support a shorter distance scale, or high H_0 . The same argument applies to the fact that we can resolve the brightest stars in galaxies in the direction of the Virgo cluster^{30,31}.

Other promising techniques may eventually provide a way of bypassing entirely the laborious steps of the traditional distance ladder. Time delays (δt) between images of gravitationally lensed quasars^{32,33} give a direct physical estimate of the scale of the Universe: $\delta t \approx \theta^2 H_0^{-1}$, where θ is the angle by which the lensed images are separated. But current estimates from the best studied example, Q0957+561, are close to both of the distance ladder estimates, giving a range of values consistent with high or low H_0 depending on model parameters³⁴. Observations of comptonization of background radiation by X-ray emitting gas in one specific galaxy cluster³⁵ currently suggest low H_0 , based on a simple smooth spherical model for the cluster gas. A reliable independent estimate of H_0 from this technique will require a large statistical sample to control effects of cluster elongation and selection effects in cluster samples.

We have raised several issues that seem to be particularly critical. Is SN1937C a fair representative of the class of type Ia supernovae? If so, are the nearby calibrator galaxies for TF, SBF and PNLF indeed unrepresentative of more distant galaxies, by an offset that happens to be the same for three independent kinds of observation? Must we modify the standard model for type Ia supernovae, or the apparently successful model for the physics of the expanding photospheres of type II supernovae? The issues are sharply defined, and they are likely to be resolved by observational programs that are either in progress or at least seem feasible. An important test, now underway using the Hubble Space Telescope, is the measurement of additional Cepheid distances to obtain type Ia, PNLF, SBF, TF and EPM calibrations in other, more distant galaxies; in addition, a valuable check on EPM systematic errors will be provided by the fortuitous appearance this year of a new type II supernova, 1993J, in M81, which now also has a firm Cepheid distance. $\square$

M. Fukugita is at the Yukawa Institute of Kyoto University, Kyoto 606, Japan and at the Institute of Advanced Study, Princeton, New Jersey 08540, USA. C. J. Hogan is at the University of Washington, Seattle, Washington 98195, USA. P. J. E. Peebles is at Princeton University, Princeton, New Jersey 08544, USA.

- Sandage, A. *Astrophys. J.* **402**, 3–14 (1993).
- de Vaucouleurs, G. in *10th Texas Symp. on Relativistic Astrophysics* (eds Ramaty, R. & Jones, F. C.) *Ann. N.Y. Acad. Sci.* **375**, 90–122 (1981).
- Schmidt, B. P., Kirshner, R. P. & Eastman, R. G. *Astrophys. J.* **395**, 366–386 (1992).
- Panagia, N. et al. *Astrophys. J.* **380**, L23–L26 (1991).
- Freedman, W. & Madore, B. in *New Perspectives on Stellar Pulsation and Pulsating Stars*, IAU Colloq. No. 139 (eds Nemec, J. & Matthews, J.) (Cambridge Univ. Press, 1993).
- Freedman, W. et al. *Astrophys. J.* (submitted).
- Kristian, J., Sandage, A. & Westphal, J. *Astrophys. J.* **221**, 383–394 (1978).
- Lauer, T. & Postman, M. *Astrophys. J.* **400**, L47–L50 (1992).
- Mould, J. R. et al. *Astrophys. J.* **409**, 14–27 (1993).
- Sandage, A. & Tammann, G. A. *Astrophys. J.* **365**, 1–12 (1990).
- Leibundgut, B. & Pinto, P. A. *Astrophys. J.* **401**, 49–59 (1992).
- Branch, D. & Miller, D. L. *Astrophys. J.* **405**, L5–L8 (1993).
- Sandage, A., Saha, A., Tammann, G. A., Panagia, N. & Macchetto, D. *Astrophys. J.* **401**, L7–L10 (1992).
- Arnett, W. D., Branch, D. & Wheeler, J. C. *Nature* **314**, 337–338 (1985).
- Branch, D. *Astrophys. J.* **392**, 35–40 (1992).
- Khokhlov, A., Muller, E. & Höflich, P. *Astr. Astrophys.* **270**, 223–248 (1993).
- Yamaoka, H., Shigeyama, T. & Nomoto, K. Univ. of Tokyo preprint (1993).
- Jacoby, G. H. et al. *Publ. astr. Soc. Pacif.* **104**, 599–662 (1992).
- van den Bergh, S. *Publ. astr. Soc. Pacif.* **104**, 861–883 (1992).
- Tully, R. B. & Fisher, J. R. *Astr. Astrophys.* **54**, 661–673 (1977).
- Fukugita, M., Okamura, S. & Yasuda, N. *Astrophys. J.* **412**, L13–L16 (1993).
- Ichikawa, T. & Fukugita, M. *Astrophys. J.* **394**, 61–80 (1992).
- Tony, J. *Astrophys. J.* **373**, L1–L4 (1991).
- Jacoby, G. H., Ciardullo, R. & Ford, H. C. *Astrophys. J.* **356**, 332–349 (1990).
- Madore, B. F. & Freedman, W. L. *Publ. astr. Soc. Pacif.* **103**, 933–957 (1991).
- Pierce, M. *Astrophys. J.* (in the press); National Optical Astronomy Observatories preprint No. 502.
- Maoz, D. et al. *Astrophys. J.* **402**, 69–75 (1993).
- Huchra, J. P. *Science* **256**, 321–325 (1992).
- Pierce, M. et al. in *New Perspectives on Stellar Pulsation and Pulsating Stars*, IAU Colloq. No. 139 (eds Nemec, J. & Matthews, J.) (Cambridge Univ. Press, 1993). (See ref. 5)
- Shanks, T. et al. *Mon. Not. R. astr. Soc.* **256**, 29P–32P (1992).
- Pierce, M. J., McClure, R. D. & Racine, R. *Astrophys. J.* **393**, 523–529 (1992).
- Press, W. H., Rybicki, G. B. & Hewitt, J. N. *Astrophys. J.* **385**, 416–420 (1992).
- Lehar, J., Hewitt, J. N., Roberts, D. H. & Burke, B. F. *Astrophys. J.* **384**, 453–466 (1992).
- Bernstein, G. M., Tyson, J. A. & Kochanek, C. S. *Astro. J.* **105**, 816–830 (1993).
- Birkinshaw, M., Hughes, J. P. & Arnaud, K. A. *Astrophys. J.* **379**, 466–481 (1992).
- Peebles, P. J. E. *Principle of Physical Cosmology* (Princeton Univ. Press, 1993).

ACKNOWLEDGEMENTS. This work was inspired by the January 1993 Aspen Winter Conference on Astrophysics; we thank the participants for their guidance. We are grateful to D. Branch, G. Jacoby, R. Kirshner, J. Peacock, M. Pierce, S. van den Bergh, J. Tony and C. Wheeler who helped improve (while not necessarily approving) a draft version of the report. This work was supported (in part) at Kyoto by the Japan-US Cooperative Science Programme, by JSPS and the US NSF; at the University of Washington by NASA and the US NSF; and at Princeton by the US NSF.

The yeast *DOA4* gene encodes a deubiquitinating enzyme related to a product of the human *tre-2* oncogene

Feroz R. Papa & Mark Hochstrasser

Department of Biochemistry and Molecular Biology, The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA

Modification of specific intracellular proteins by ubiquitin targets them for degradation. We describe a yeast enzyme, Doa4, that is integral to the degradation of ubiquitinated proteins and is required in diverse physiological processes. Doa4 appears to function late in the proteolytic pathway by cleaving ubiquitin from substrate remnants still bound to protease. The human *tre-2* oncogene encodes a deubiquitinating enzyme similar to Doa4, indicating a role for the ubiquitin system in mammalian growth control.

FOR many short-lived eukaryotic proteins, covalent attachment to the polypeptide ubiquitin is a prerequisite for their degradation; examples include the destruction of cyclins in cell cycle control¹ and the degradation of proteins that regulate development, such as the vertebrate *c-mos* protein kinase² or the yeast MAT α 2 transcription factor^{3,4}. The ubiquitin system is essential for phenomena as diverse as the heat-shock response and DNA repair, and both ubiquitin and the enzymes involved in ubiquitin metabolism are highly conserved among eukaryotic phyla⁵⁻⁹. Using a MAT α 2 derivative as substrate, we previously identified a ubiquitin-dependent degradation pathway in the yeast *Saccharomyces cerevisiae*, the DOA (for degradation of alpha) pathway^{4,10}. This pathway, which requires two ubiquitin-conjugating (Ubc) enzymes, Ubc6 (Doa2) and Ubc7, targets α 2 through an element within the first 67 residues of the α 2 repressor, the *Deg1* degradation signal. Many more gene products have been implicated in α 2 degradation, based on genetic analysis⁴.

Here we show that one component of the DOA pathway, Doa4, is a deubiquitinating enzyme. Deubiquitinating enzymes serve a number of functions^{7,11}. First, ubiquitin must be cleaved from a set of biosynthetic precursors, which occur either as a series of ubiquitin monomers in head-to-tail linkage or as fusions to certain ribosomal proteins⁵. Second, ubiquitin must be recycled from intracellular conjugates, both to maintain adequate pools of free ubiquitin and, in principle at least, to reverse the modification of inappropriately targeted proteins. Finally, deubiquitinating reactions may be integral to the degradation of ubiquitinated proteins by the 26S proteasome, a complex ATP-dependent enzyme whose exact composition and range of activities remain poorly characterized^{6,12-14}. Consistent with these diverse functions, eukaryotic cells have been shown to contain many distinct deubiquitinating enzymes^{11,15-17}. Previously, four deubiquitinating enzyme genes, *YUH1*, *UBP1*, *UBP2* and *UBP3*, had been identified in *S. cerevisiae*¹⁷⁻¹⁹. A mutant strain lacking all four genes grows normally and shows continued high deubiquitinating activity¹⁷.

We demonstrate here that Doa4 is a member of what appears to be a large family of deubiquitinating enzymes found in yeast, plants and animals. Elimination of Doa4 from yeast cells results in poor growth and many additional defects. Proteolysis of all tested ubiquitin-dependent substrates is strongly inhibited in *doa4* mutants. Our data suggest that the Doa4 enzyme functions late in the proteolytic pathway in conjunction with (or as part of) the 26S proteasome. Doa4 is more similar in sequence to a

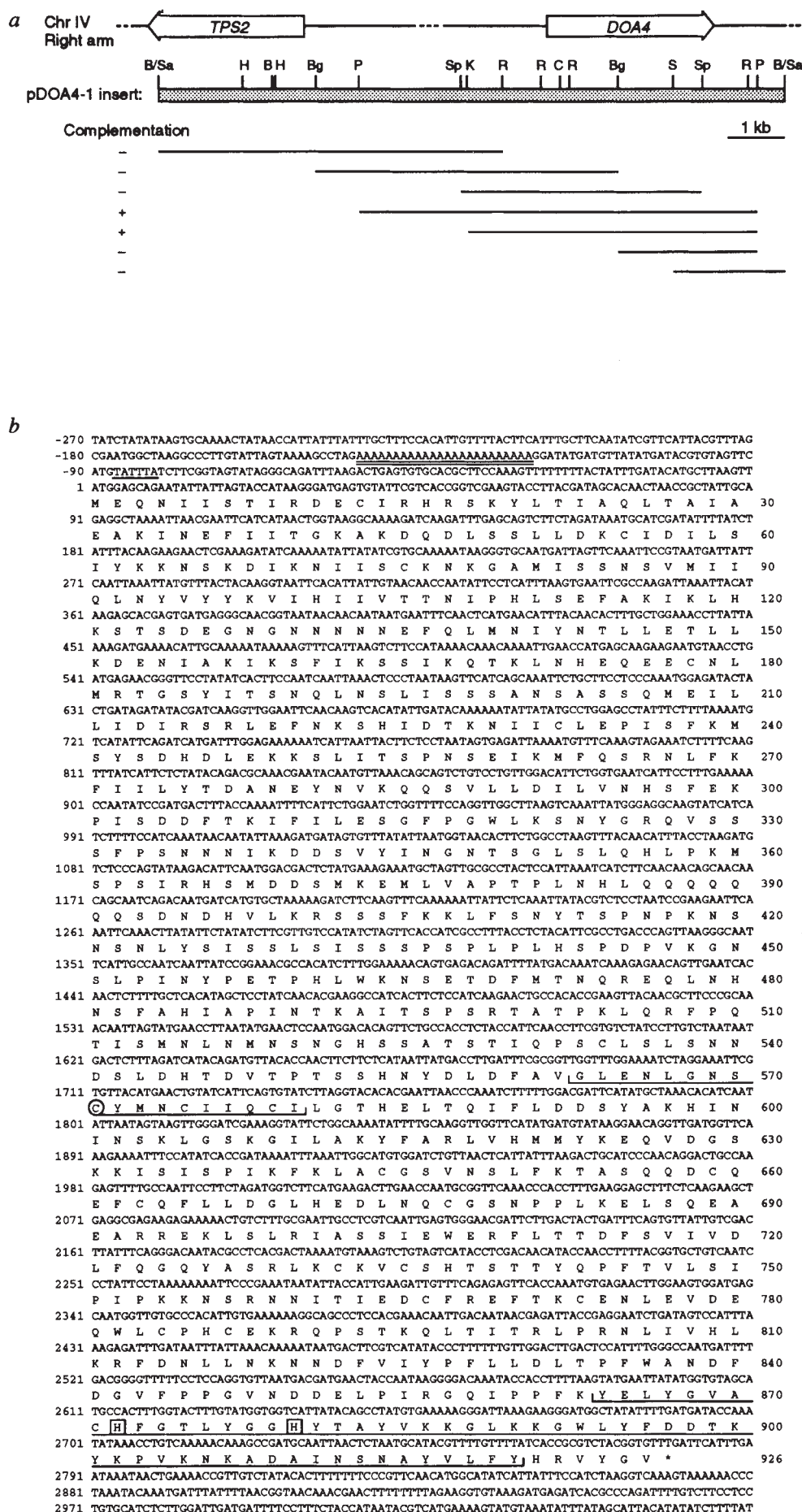
product of the human *tre-2* oncogene²⁰ than to any of the known yeast Ubp proteins, and we show that *tre-2*, whose function was previously unknown, also encodes a deubiquitinating enzyme. Moreover, a mutation in *DOA4* has properties that parallel those of an oncogenic variant of *tre-2*. These findings suggest that perturbation of ubiquitin-dependent proteolysis in mammalian cells can lead to tumorigenic growth.

DOA4 encodes a deubiquitinating enzyme

The *DOA4* gene was isolated by its ability to restore degradation of a MAT α 2- β -galactosidase fusion protein, *Deg1*- β gal, in the degradation-defective *doa4-1* mutant¹⁰ (Fig. 1a). The nucleotide sequence of the complementing region of the cloned DNA (Fig. 1b) has a single long potential open reading frame (ORF), which would encode a 926-residue (*M*, 105K) protein. A search of the sequence databases revealed two short regions of similarity between Doa4 and three ubiquitin-specific processing enzymes from yeast, Ubp1-Ubp3 (refs 17, 19) (Fig. 2d, e). These proteins had been identified on the basis of their ability to cleave ubiquitin from linear ubiquitin-protein fusions, including natural ubiquitin precursors. The Ubp1-Ubp3 enzymes are dissimilar to each other except in the two stretches also shared by Doa4.

To determine whether Doa4 has protein-deubiquitinating activity, the *DOA4* gene and a gene encoding Ub-Met- β gal, in which ubiquitin is fused to the N-terminus of β -galactosidase⁹, were coexpressed in *Escherichia coli* (which lacks a ubiquitin system). As shown by immunoblot analysis (Fig. 3a), Doa4 cleaved Ub-Met- β gal to Met- β gal (lane 3) to an extent comparable to that observed with Ubp1 (lane 2). When the Met residue in Ub-Met- β gal was replaced by Leu, cleavage was unaffected, but very little Leu- β gal product accumulated (Fig. 3a, lane 6), as would be predicted if the Doa4-catalysed scission occurred precisely at the junction between ubiquitin and Leu- β gal because Leu- β gal, unlike Met- β gal, is rapidly degraded in *E. coli*⁹. Deubiquitinating enzymes have properties of thiol proteases¹¹. The sequence elements shown in Fig. 2d, e include absolutely conserved Cys and His residues. The codon for Cys 571 of Doa4 was changed to a Ser or Ala codon, and the resulting Doa4 mutants, and the corresponding glutathione *S*-transferase (GST)-Doa4 fusion proteins, were tested for deubiquitinating activity. The *doa4*^{Ser571} and *doa4*^{Ala571} proteins, as well as their GST derivatives, were unable to process the Ub-Met- β gal substrate (Fig. 3a, lanes 4 and 5, 9 and 10). Moreover, neither *doa4*^{Ser571} nor *doa4*^{Ala571} was functional in yeast cells (Fig. 3d). Thus, Doa4 can specifically cleave ubiquitin from linear ubiquit-

FIG. 1 The *DOA4* gene. **a**, Restriction map of a ~11 kb genomic fragment containing *DOA4* and identification of DNA subclones that complement the *doa4-1* proteolytic defect. +, rescue of the *Deg1*- β gal degradation defect; -, failure to reverse the defect. Restriction enzyme sites: B, *Bam*HI; Bg, *Bgl*II; C, *Clal*; H, *Hind*III; K, *Kpn*I; P, *Pst*I; R, *Eco*RV; S, *Sal*I; Sa, *Sau*3AI. **b**, Nucleotide sequence of *DOA4* and predicted amino acid sequence of the Doa4 protein. The conserved domains that include a Cys (circled) and two His (boxed) residues (Fig. 2d, e) are bracketed. Upstream of the ORF, two sequences that may play a role in transcription initiation are highlighted, a poly(dA) tract (double underline)³¹ and a potential TATA element (underline)³². **METHODS.** The *DOA4* gene was cloned from a yeast genomic library³³ by complementation of *doa4-1* mutant MH11D5-8a (see Fig. 4). Fragments from plasmid pDOA4-1, which contained the smallest complementing DNA insert, were subcloned into YCplac33 (ref. 34). Complementation activity was localized to a 5.4-kb *Kpn*I-*Pst*I fragment (plasmid pDOA4-8). That the complementing gene was indeed *DOA4* was verified by showing close linkage of an integrated fragment of the cloned DNA, marked with *URA3*, and the original *doa4-1* mutation (12 tetrads, all parental ditypes). DNA was sequenced on both strands³⁵. Although 24 oligo(dA) residues in the poly(dA) tract are shown, the number sequenced ranged between 23 and 25 in different M13 subclones. The DNA sequence of *DOA4* has been deposited in GenBank (accession no. U02518). The codon adaptation index³⁶ of *DOA4* is 0.154, which is within the range expected of weakly expressed yeast genes. There is an unpublished yeast sequence, SSV7, in the GenBank database (L08070) that encodes a protein 97% identical to Doa4. The two genes are likely to be the same, with differences due to SSV7 sequencing errors or polymorphisms. The 5' sequence of the pDOA4-1 insert matched a region of the *TPS2* gene on chromosome IV (1.6 kb overlap)³⁷.



in-protein fusions, and this activity, as well as Doa4 function in yeast, requires the conserved Cys residue.

Ubiquitin-dependent proteolytic substrates require attachment of an isopeptide-linked multiubiquitin chain for efficient degradation²¹. The ability of Doa4 to deubiquitinate a protein with ubiquitin in isopeptide linkage was tested with a ubiquitin-Lys48-ubiquitin dimer²² as substrate, in which the α -carboxyl group of one ubiquitin is attached to the ϵ -amino group of Lys 48 in the second ubiquitin (the same linkage found in the multiubiquitin chains formed on proteolytic substrates²¹). GST-Doa4 was isolated from *E. coli* and incubated with ubiquitin dimer. The bulk of the dimer molecules were cleaved to free ubiquitin (Fig. 3c, lane 4). That this activity was due to GST-Doa4, rather than a copurifying *E. coli* enzyme, was demonstrated by the inability of GST-doa4^{Ser571} to cleave the substrate (Fig. 3c, lane 3). Thus, Doa4 is able to cleave both peptide-

linked and isopeptide-linked ubiquitin moieties from substrates.

Inhibition of proteolysis in *doa4* mutants

To assess the phenotype of a *doa4* null mutant, the *DOA4* gene was deleted (Fig. 4a). Phenotypically, *doa4Δ* cells were indistinguishable from the original *doa4-1* mutant. *Doa4*⁻ cells had multiple defects, with aberrations ranging from slow growth to a defect in DNA repair (based on sensitivity to ultraviolet or γ -radiation) (Fig. 4c). Thus, the Doa4 deubiquitinating enzyme is required in diverse physiological processes. On the other hand, the only condition that appeared to increase *DOA4* transcript levels was entry into stationary phase (Fig. 4b). Expression of several other ubiquitin system genes is elevated in stationary phase⁵.

The abnormalities of *doa4* cells may reflect defects in ubiquitin-dependent proteolysis inasmuch as mutations in *UBC* or protea-

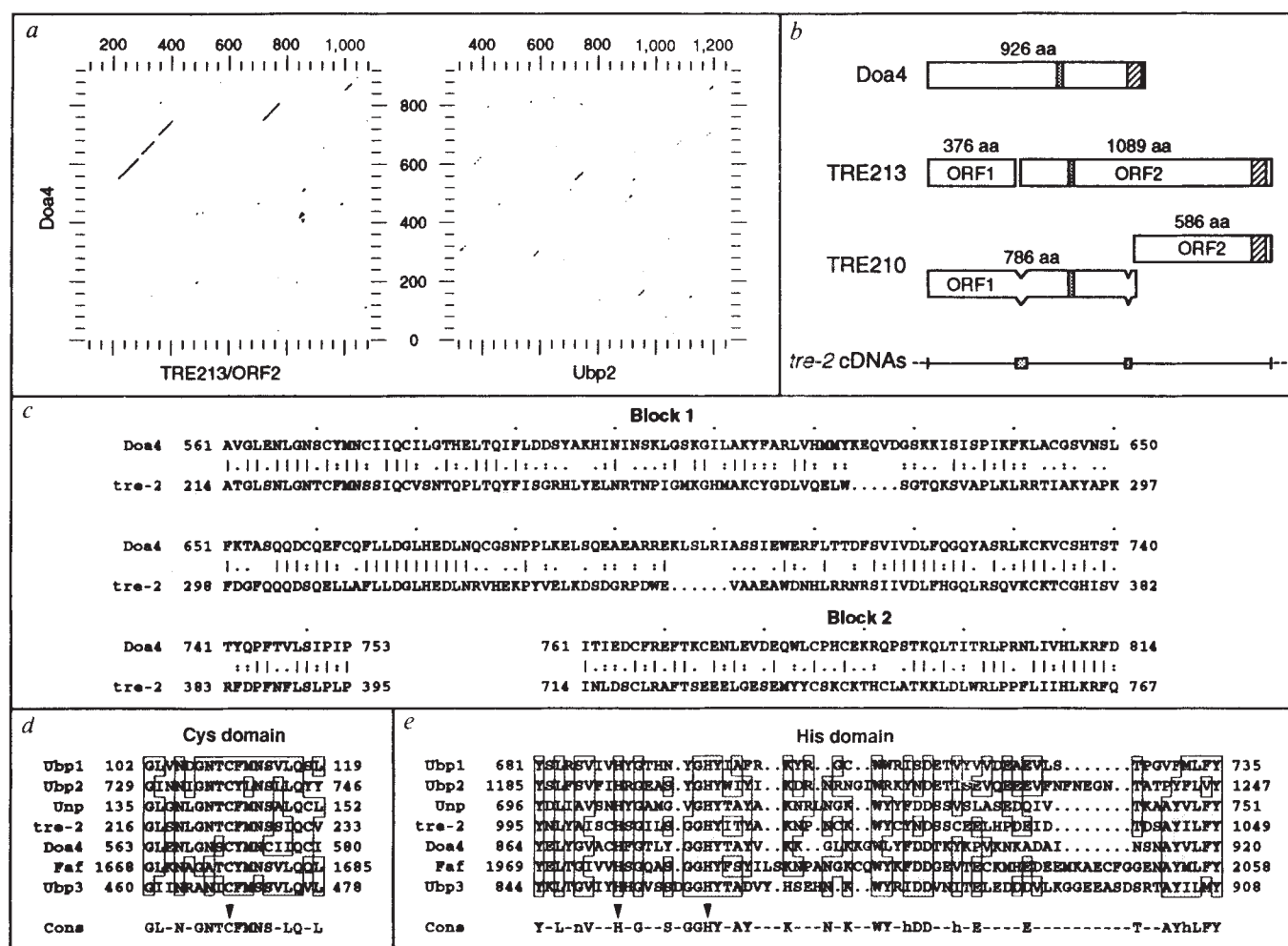


FIG. 2 The deubiquitinating enzyme family. *a*, DOTplots of Doa4 versus human *tre-2* and yeast Ubp2. Window, 30; stringency, 16. *b*, The open reading frames of yeast *DOA4* and human *tre-2* (ref. 20) genes, the latter predicted from two cDNA clones, which differ by two segments present in the TRE213 clone but absent from TRE210 (boxed in cDNA; indents in TRE210/ORF1). The Cys (stippled) and His (cross-hatched) domains are highlighted. aa, Amino acids. *c*, Sequence conservation between two long elements in Doa4 and *tre-2* (TRE213/ORF2). The more N-terminal region, block 1 (39% identity/60% similarity³⁸), includes the Cys domain (*d*); the smaller element is called block 2 (43%/61%). A third region, block 3, is referred to as the His domain (*e*). Weaker sequence similarities between various family members was observed in blocks 1 and 2, for example the region corresponding to residues 656–675 of Doa4. *d* and *e*, Conserved elements in Doa4 and

other deubiquitinating enzymes around the presumptive active site Cys and His residues (arrowheads), including consensus sequences (≥ 4 identities). Residues are boxed if present in at least three of the proteins or, in the case of small uncharged (n) or hydrophobic (h) residues, share the indicated property in at least six of the proteins. Sequences compared to Doa4 are yeast Ubp1–Ubp3 (refs 17, 19); mouse Unp²⁴; human *tre-2* (TRE213/ORF2)²⁰; and *Drosophila fat facets* (*faf*)²⁵ (a 22-residue segment was deleted after residue 2,029 of *faf*). Sequences were aligned initially with PILEUP. Many partial cDNA sequences in the GenBank database, identified with BLAST³⁹, also show significant similarity to regions within blocks 1–3; these include human (T04867, T09031, Z12710, Z15701, Z21167), *Caenorhabditis elegans* (M80041, Z14720, Z14811), *Arabidopsis thaliana* (Z17750, Z25610), and canine (L03387) sequences.

somal genes result in many of the same defects⁸. Figure 5a and b shows the defect of the *doa4-1* mutant in the degradation of *Deg1*- β gal and of intact $\alpha 2$ repressor. Inactivation of Doa4 resulted in a 4–5-fold increase in the half-life of $\alpha 2$ at 30° C (Fig. 5b), whereas *Deg1*- β gal degradation was inhibited at least 10-fold relative to wild-type cells (Fig. 5a). Two well-characterized classes of artificial substrates of the ubiquitin-dependent proteolytic system in yeast have been described⁹. One class, the N-end-rule substrates, requires Ubc2 for multiubiquitination, whereas the other, represented by the Ub-Pro- β gal protein, depends on a distinct Ubc enzyme, Ubc4. Degradation of Leu- β gal, an N-end-rule substrate, and Ub-Pro- β gal was inhibited ~10-fold and ~20-fold, respectively, in the *doa4Δ* mutant (Fig. 5c, d). Multiubiquitinated forms of both Leu- β gal and Ub-Pro- β gal were present but at reduced levels in *doa4Δ* cells relative to wild type. Based on the strong inhibitory effects on the turnover of these substrates, which are subject to distinct ubiquitination mechanisms, we conclude that Doa4 plays a key role in the ubiquitin-dependent protein degradation system of *S. cerevisiae*.

If Doa4 functioned *in vivo* to deubiquitinate proteins before their targeting to the 26S proteasome, raising intracellular levels of Doa4 might also be expected to inhibit substrate proteolysis. Overexpression of Ubp2, for instance, inhibits Ub-Pro- β gal degradation (see ref. 17). *DOA4* was subcloned into a high-copy 2 μ -based plasmid that was introduced into wild-type yeast cells. Levels of *DOA4* mRNA in these cells were increased at least 20-fold above normal (Fig. 4b, lanes 1, 2). Surprisingly, elevated expression of Doa4 accelerated, rather than inhibited, *Deg1*- β gal and $\alpha 2$ degradation (Fig. 5e, and data not shown). The turnover of $\alpha 2$ increased ~30–40% relative to cells bearing the control plasmid (Fig. 5e). Under certain conditions, Ub-Pro- β gal turnover was also increased slightly by overproduction of Doa4 (data not shown). Thus, intracellular levels of Doa4 can be at least partially rate-limiting for ubiquitin-dependent protein degradation.

A dominant negative mutation in *DOA4*

Expression of the Doa4 active-site mutant, *doa4*^{Ser571}, from a high-copy plasmid in wild-type cells inhibited degradation of all

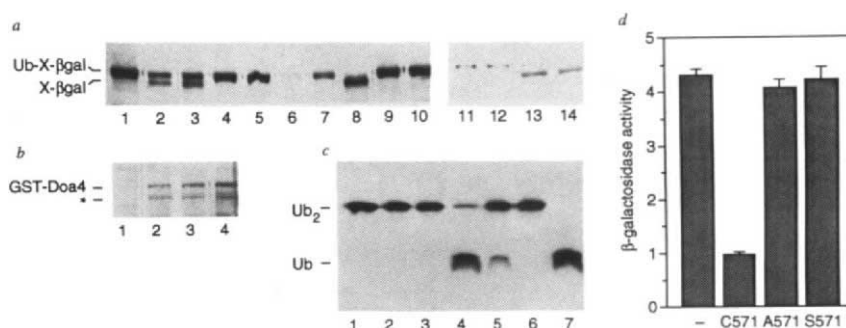
ubiquitin-dependent substrates tested (Fig. 5e, f). Overproduction of *doa4*^{Ser571} also caused cellular defects comparable to those resulting from deletion of *DOA4*, such as slower growth. High-copy *doa4*^{Ser571} expression resulted in a ~30% increase in the doubling time of wild-type MHY102 cells in minimal medium. That these abnormalities resulted from expression of *doa4*^{Ser571} protein (rather than competition of the high-copy *doa4*^{Ser571} gene or messenger RNA for a factor(s) required for expression of endogenous *DOA4*) was ascertained by inserting two extra nucleotides near the 5' end of the *doa4*^{Ser571} ORF (Fig. 5 legend). The resulting frameshift mutation completely eliminated the defect associated with high-copy expression of *doa4*^{Ser571} (data not shown). We conclude that the dominant-negative defect of *doa4*^{Ser571} resulted from elevated expression of a catalytically inactive protein. This could be explained as either a titration of substrate by the inactive protein, formation of an inactive Doa4 heteromer, or competition with normal Doa4 for binding to a component(s) of the proteolytic machinery, perhaps the 26S protease itself. The first possibility seems improbable, given the abundance of ubiquitin pathway substrates *in vivo* and the fact that *doa4*^{Ser571}-substrate interactions would be self-limiting.

Defective metabolism of ubiquitin conjugates

To gain insight into the mechanistic role(s) of the Doa4 enzyme in ubiquitin-dependent proteolysis, we examined intracellular ubiquitin and ubiquitin-protein conjugates by anti-ubiquitin immunoblot analysis (Fig. 6a). Only a small decrease in free ubiquitin levels was observed in *doa4* cells compared with congenic wild-type cells (Fig. 6a, lanes 2–5). Moreover, expression of a ubiquitin gene from a strong promoter, which increases free ubiquitin levels ~100-fold²³, did not compensate for loss of Doa4, as measured by either growth rate or Ub-Pro- β gal degradation (data not shown). The most striking difference between *DOA*⁺ and *doa4* cells apparent from Fig. 6a was the accumulation of relatively small ubiquitin-containing species in the mutant (arrowhead and bracket). The smallest of these (arrowhead) was only slightly larger than ubiquitin itself; assuming this is a ubiquitin-oligopeptide conjugate, the peptide(s) is unlikely to exceed ~8–15 residues, based on comparisons with peptide-

FIG. 3 Deubiquitination reactions catalysed by Doa4 and *tre-2*.

a, Deubiquitination of ubiquitin- β -galactosidase fusion proteins expressed in bacteria. Substrate was Ub-Met- β gal in lanes 1–5 and 7–14 and Ub-Leu- β gal in lane 6. Co-expressed plasmids were (lanes): 1, YCplac33; 2, pJT60 (*UBP1*); 3, pDOA4-8; 4, pDOA4-8^{Ala571}; 5, pDOA4-8^{Ser571}; 6, pDOA4-8; 7, pGEX-KG; 8, pGEX-DOA4; 9, pGEX-DOA4^{Ala571}; 10, pGEX-DOA4^{Ser571}; 11, pGEX-KG; 12, pGEX-TRE210/ORF1; 13, pGEX-TRE213/ORF2; and 14, pGEX-TRE210/ORF1-TRE213/ORF2. **b**, Comparable levels of GST-Doa4 and GST-*doa4* point mutants in bacteria, as measured by anti-GST immunoblot analysis. Cells carried the following plasmids: for lane 1, pGEX-KG; lane 2, pGEX-DOA4; lane 3, pGEX-DOA4^{Ala571}; and lane 4, pGEX-DOA4^{Ser571}. Asterisk indicates a GST-Doa4 fragment. Full-length GST-TRE210/ORF1 and GST-TRE213/ORF2 proteins were both present at lower but detectable levels (data not shown). **c**, Liberation of ubiquitin (Ub) from an isopeptide-linked ubiquitin dimer (Ub₂) by GST-Doa4 and GST-*tre-2* proteins partially purified from *E. coli*. Lanes 1 and 7, untreated Ub₂ and ubiquitin, respectively. Protein added: lane 2, GST; lane 3, GST-*doa4*^{Ser571}; lane 4, GST-*doa4*; lane 5, GST-TRE213/ORF2; and lane 6, GST-TRE210/ORF1. **d**, Inactivity of *doa4*^{Ser571} and *doa4*^{Ala571} in yeast. *Deg1*- β gal activity was measured in mutant MHY102 cells bearing the following plasmids: YCplac33 (–); pDOA4-8 (C571); pDOA4-8^{Ala571} (A571); pDOA4-8^{Ser571} (S571). Values are the average of ≥3 measurements from 2–3 transformants; error bars indicate standard deviations. The *doa4*^{Ser571} and *doa4*^{Ala571} plasmids also fail to correct the *doa4* growth and sporulation defects. **METHODS**. The *doa4*^{Ala571} and *doa4*^{Ser571} alleles were constructed by oligonucleotide-directed mutagenesis³⁵. GST fusions were made by



amplification of *DOA4* or *tre-2* sequences by PCR (done in duplicate) and insertion into pGEX-KG³⁵ downstream of the glutathione-S-transferase coding element. Ub-Met- β gal and Ub-Leu- β gal were expressed from pACYC184-based plasmids. Plasmid-bearing *E. coli* MC1061 cells were analysed by immunoblotting with anti- β gal (Cappel). *Doa4*^{Ser571} and *doa4*^{Ala571} also did not deubiquitinate Ub-Leu- β gal, and the level of Ub-Leu- β gal in these experiments was comparable to that of Ub-Met- β gal (data not shown); the reduction of unprocessed Ub-Leu- β gal in lane 6 may reflect a mass action effect due to Leu- β gal degradation. Ub-X- β gal molecules are stable in *E. coli*⁴⁰. Antibody to *S. japonicum* GST (Amrad) was used in **b**. Assays of Ub₂ cleavage (**c**) were done with GST-Doa4 or GST-*tre-2* proteins partially purified from *E. coli* JM101 (ref. 35). Equal amounts of protein were incubated at 37 °C with 60 ng Ub₂ (ref. 22) in 50 mM Tris-HCl, pH 7.6, 4 mM DTT. Proteins were detected by anti-ubiquitin (East Acres Biologicals) immunoblot analysis with the ECL system (Amersham). β -galactosidase activity (**d**) was measured as described¹⁰.

tagged ubiquitin derivatives^{3,23}. Similarly, the electrophoretic mobilities of the ubiquitin-reactive species (bracketed) suggested that they may represent small multiubiquitin chains (dimers, trimers, and so on) linked to a heterogeneous set of oligopeptides. Overexpression of Doa4 further reduced the low levels of these ubiquitin species in wild-type cells (Fig. 6a, lanes 1 and 2), indicating that the rate of metabolism of these species was limited by intracellular Doa4 levels.

tre-2 also encodes a deubiquitinating enzyme

The strongest similarity between Doa4 and proteins in the current databases was found with a set of proteins predicted to be encoded by a human oncogene, *tre-2* (ref. 20). The *tre-2* gene gives rise to a least two alternatively spliced mRNAs, both encoding two ORFs²⁰ (Fig. 2b). The similarity between the yeast and human proteins extends over several sequence blocks that comprise over 300 residues (Fig. 2a, c). Inasmuch as the yeast Ubp1, Ubp2, Ubp3 and Doa4 proteins are dissimilar to each other except in the two short stretches already mentioned, it is striking that the Doa4 and human *tre-2* proteins show extensive similarity, suggesting that they may represent a functionally distinct class of deubiquitinating enzymes. A putative mouse proto-oncoprotein, Unp, has been described recently²⁴ that shows strong similarity to *tre-2* and to Doa4 as well. Doa4 is also related to a *Drosophila* protein involved in eye and embryonic development, *fat facets* (*faf*)²⁵ (Fig. 2).

To determine whether the close sequence similarity of Doa4 and *tre-2* reflects a similarity in enzymatic function, deubiquitinating activity of several *tre-2* ORFs was measured. A GST-TRE213/ORF2 protein had catalytic activities comparable to Doa4, whereas GST-TRE210/ORF1 was inactive (Fig. 3a, lanes 12 and 13; Fig. 3c, lanes 5 and 6). The lack of enzyme activity of GST-TRE210/ORF1 is consistent with the absence of the highly conserved His domain from this polypeptide (Fig. 2b). The TRE210/ORF1 protein was the only polypeptide found to be tumorigenic when individual ORFs derived from *tre-2* were expressed at high levels in nude mice²⁰. Tumorigenicity was suppressed if TRE213/ORF2 was fused to TRE210/ORF1 to create

the long ORF that would be formed in a *tre-2* mRNA bearing only the second of the two unique TRE213 segments (Fig. 2b). Correspondingly, we found that this 'full-length' protein also had deubiquitinating activity (Fig. 3a, lane 14).

Discussion

Our results show that Doa4 is a central component of the yeast ubiquitin-dependent proteolytic system. Deletion of the *DOA4* gene leads to multiple abnormalities (Fig. 4c), despite the fact that *S. cerevisiae* cells have at least four additional deubiquitinating enzymes¹⁷. (In the nomenclature of the previously described yeast deubiquitinating enzymes, Doa4 would be called Ubp4, the fourth identified member of the Ubp class.) The functional diversity of deubiquitinating enzymes implied by our *in vivo* data may reflect differences in their substrate specificity or in their mechanistic roles within the ubiquitin pathway. *A priori*, Doa4/Ubp4-catalysed deubiquitination reactions could be important in ubiquitin precursor processing, in removing ubiquitin from substrates before their commitment to proteolysis, and/or in the proteolytic phase of the degradation pathway. Activity assays with isolated Doa4 (Fig. 3) are consistent with all three roles. However, the data in Figs 4–6 strongly suggest that Doa4 functions primarily, if not exclusively, in the late stages of proteolysis *in vivo*.

A critical role for Doa4 in ubiquitin precursor processing is inconsistent with several results. First, processing of a model ubiquitin-protein fusion, Ub-Leu- β gal, to Leu- β gal is unimpaired in *doa4* cells (Fig. 5c). Second, immunoblot analysis of *doa4* mutants (Fig. 6a) does not show accumulation of a species of the size expected for the Ubi1 and Ubi2 ubiquitin precursors²⁶. Finally, there is only a minor decrease in free ubiquitin levels in *doa4* relative to wild-type cells (Fig. 6a), and the mutant phenotype of a *doa4* strain is not suppressed by overexpression of ubiquitin. These data therefore render unlikely any model that invokes a decrease in ubiquitin availability to explain the effect of *doa4* mutations on ubiquitin-dependent proteolysis. Such models would include, for example, those that posit a primary role for Doa4 in recycling of ubiquitin from aberrant ubiquitin adducts

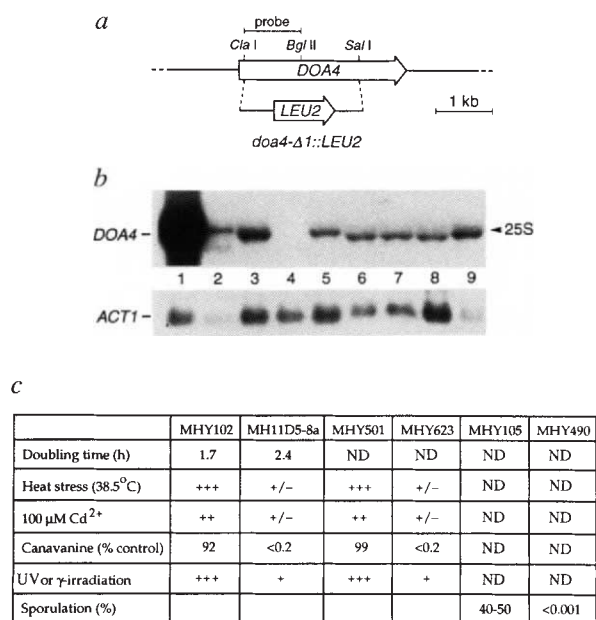


FIG. 4 a, Structure of the *doa4-Δ1::LEU2* null allele. b, Northern RNA hybridization analysis of *DOA4* gene expression. Lane 1, wild-type (DBY1705) cells carrying YEplacDOA4; lane 2, DBY1705 carrying the YEplac195 vector; lane 3, *doa4-1* (MH11D5-8a); lane 4, *doa4Δ* (MHY622); lane 5, wild type (MHY501); lane 6, MHY501 shifted to 38.5 °C for 30 min; lane 7, MHY501 treated with 100 μM cadmium acetate for 30 min; lane 8, MHY501 treated with 20 μg ml⁻¹ canavanine

or from conjugates formed with abundant intracellular nucleophiles¹¹. A primary role for Doa4 in directly altering the dynamic balance between ubiquitin addition to and removal from potential proteolytic substrates is also improbable given that elevated expression of Doa4 can accelerate, rather than

inhibit, rates of ubiquitin-dependent proteolysis *in vivo* (Fig. 5e).

Inactivation of *DOA4* leads to both a general inhibition of ubiquitin-dependent proteolysis (Fig. 5) and an accumulation of small ubiquitinated species (and a depletion of large conjugates ($M_r > 100K$); data not shown) (Fig. 6a). To account for these

FIG. 5 *doa4* mutants are impaired in the degradation of multiple ubiquitin-dependent substrates *in vivo*. **a**, Degradation of $\alpha 2$ and Deg1- β gal in wild-type (MHY102) and congenic *doa4-1* (MH11D5-7b) cells. Proteins from radiolabelled cell extracts were precipitated with an antibody to $\alpha 2$ (ref. 10). Relative molecular masses of marker proteins are indicated. **b**, Quantitation of $\alpha 2$ turnover in DOA^+ and *doa4-1* cells. The half-life of $\alpha 2$ was 4.7 min in DOA^+ and 19 min in *doa4-1* cells. **c**, Degradation of Leu- β gal, an N-end rule substrate, in *doa4 Δ* (MHY623) and DOA^+ (MHY501) cells. The open-ended bracket marks the positions of multiubiquitin-Leu- β gal conjugates. A cleavage product of M_r 90K derived from β -galactosidase is indicated⁴². Reduced amounts of the M_r 90K product were generated from Leu- β gal and especially from Ub-Pro- β gal in *doa4* cells. An anti- β gal antibody (Cappel) was used for immunoprecipitation. **d**, Degradation of Ub-Pro- β gal in MHY623 and MHY501 cells. **e**, A dominant-negative allele of *DOA4*. Degradation rates were measured in MHY102 cells carrying high-copy plasmids that expressed either Doa4 or *doa4*^{Ser571} proteins. Half-lives calculated by linear regression analysis were 5.8, 3.7 and 17 min for cells carrying YEplac195, YEplacDOA4 and YEplacDOA4^{Ser571}, respectively. **f**, Ub-Pro- β gal degradation in cells carrying high-copy pRS424 or pRSDOA4^{Ser571}.

METHODS. Pulse-chase analysis was done as described^{4,10}. The *Kpn*I-*Pst*I fragments from pDOA4-8 and pDOA4-8^{Ser571} were subcloned into either YEplac195 (ref. 34) or pRS424 (ref. 43). To verify that the dominant-negative effect of *doa4*^{Ser571} required expression of *doa4*^{Ser571} protein, the unique *Cla*I site (Fig. 1a) was cleaved, and the 2-bp 5' overhang was filled in with Klenow polymerase. Religation created an *Nru*I site and resulted in a frameshift after codon 56 of *doa4*^{Ser571}. Ub-Pro- β gal degradation in (f) was measured in MHY612 cells. MHY612 (P. Chen and M.H., unpublished) bears a *rad6 Δ ::LEU2* allele⁴⁴ in a strain otherwise isogenic with MHY501.

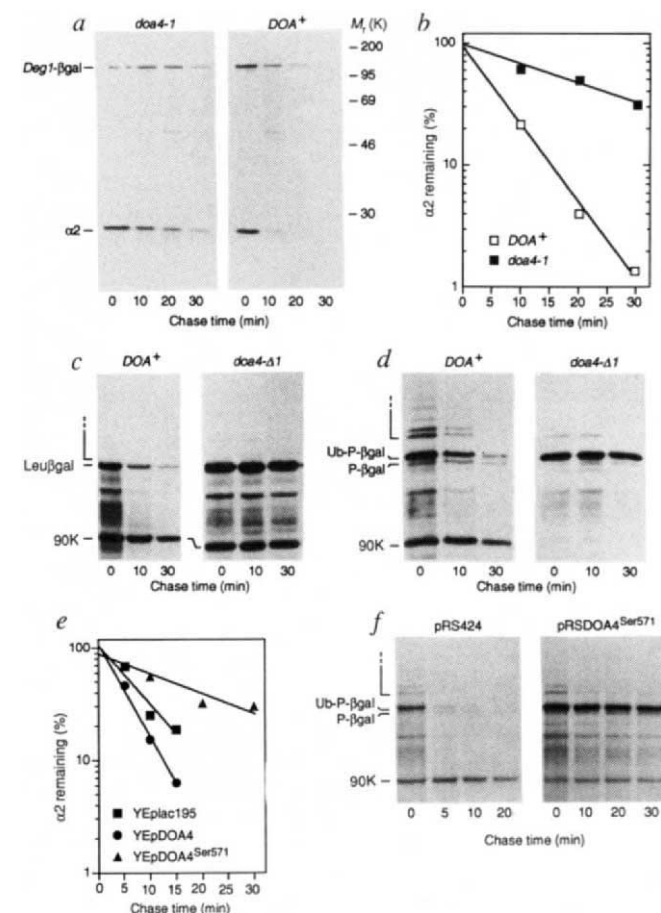
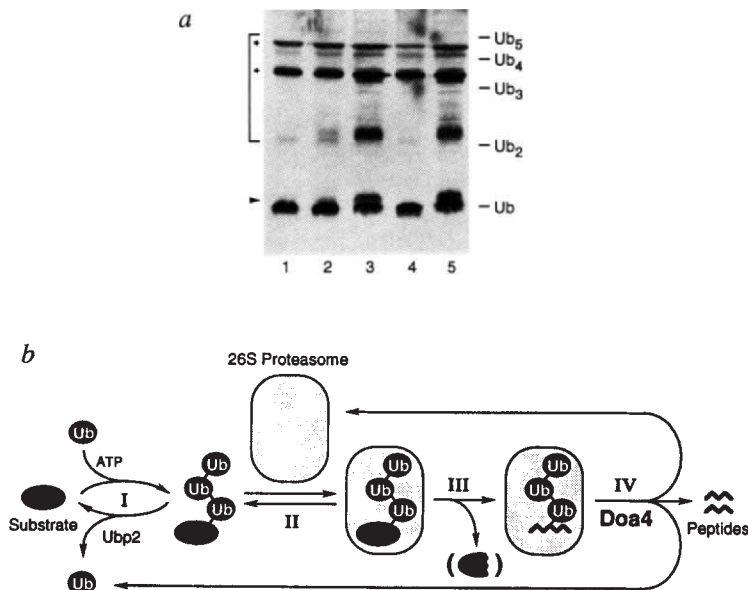


FIG. 6 **a**, Anti-ubiquitin immunoblot analysis of *doa4* and DOA^+ yeast cells and of DOA^+ cells overproducing Doa4. Lane 1, wild-type DBY1705 cells carrying the high-copy YEplacDOA4 plasmid; lane 2, DBY1705 carrying the YEplac195 vector; lane 3, *doa4-1* cells (MH11D5-8a) carrying YEplac195; lane 4, wild-type MHY501 (ref. 4); lane 5, *doa4-1 Δ* cells (MHY623). Free ubiquitin (Ub) is indicated, as is a heterogeneous set of ubiquitin-containing species larger than ubiquitin (arrowhead and bracket). Migration of unanchored multiubiquitin chains (Ub₂, Ub₃, Ub₄, Ub₅) was determined using purified Lys48-linked ubiquitin multimers (provided by S. Van Nocker). **b**, Model of Doa4 action in the ubiquitin-dependent proteolytic pathway. Doa4 is seen as the prototype of a specialized class of deubiquitinating enzymes, whose deubiquitinating activity is coupled to protein breakdown by the 26S proteasome. The possibility that relatively large cleavage products, such as the 90K product of Leu- β gal shown in Fig. 5, are released from the protease is indicated in step III.

METHODS. Exponentially growing cells in minimal medium (10 ml) were collected by centrifugation. Cells were disrupted by incubation in 50 μ l SDS loading buffer³⁵ at 100 °C for 10 min. Debris was cleared by centrifugation; ~10% of each supernatant (volumes were normalized to cell densities) was run on an 18% SDS-polyacrylamide gel; proteins were blotted onto a polyvinylidene difluoride membrane. Ubiquitin-containing proteins were detected with an anti-ubiquitin antibody (East Acres Biologicals) and the ECL system. The indicated ubiquitin-containing species were also



observed with an independent anti-ubiquitin antiserum⁴⁵ (data not shown); the two proteins marked by asterisks did not react with this antiserum.

results, we propose a model in which Doa4/Ubp4 functions by disassembling multiubiquitin chains on protein fragments still bound to the 26S proteasome (Fig. 6b). The ubiquitinated species whose levels increase in *doa4* cells cluster at positions slightly higher than those of unanchored ubiquitin monomers and oligomers (arrowhead and bracket in Fig. 6a). Such a distribution suggests that these species may be (multi)ubiquitin-peptide conjugates. By our model, such conjugates represent late intermediates in multiubiquitin-protein breakdown by the 26S proteasome. These species would persist in *doa4* cells if inhibition of their Doa4-dependent deubiquitination inhibits their disengagement from the proteasome, with proteasome-bound conjugates being inaccessible to the deubiquitinating enzymes that work on 'free' substrates (that is, substrates not bound to protease; 'I' in Fig. 6b). Slow dissociation would be expected if substrate multiubiquitin chains contribute substantially to proteasome binding. The finding that a Doa4-catalysed deubiquitination reaction(s) can limit both rates of proteolysis (Fig. 5a) and clearance of small ubiquitin conjugates (Fig. 6a) is consistent with slow conjugate dissociation. Accumulation of 26S proteasome-bound degradation intermediates would limit access of full-length multiubiquitinated substrates to proteasomes, allowing rapid disassembly of their multiubiquitin chains by deubiquitinating enzymes that work on free substrates. This interpretation requires that a significant fraction of the ubiquitinated species seen on immunoblots (Fig. 6a) represents ubiquitin conjugates still bound to 26S proteasomes. The proteasome is an abundant enzyme, constituting up to 1% of cellular protein in some cases²⁷, so detection of protease-bound substrates with limited size variation is not unlikely. (Further evidence that Doa4 and proteasome action are closely linked comes from experiments demonstrating specific genetic interactions between mutations in *DOA4* and in genes encoding proteasomal subunits (F.R.P. and M.H., unpublished results).)

The proteolytic block is incomplete in *doa4* cells, particularly for $\alpha 2$, suggesting that dissociation of ubiquitinated protein remnants from the 26S protease remains sufficiently rapid to prevent complete inactivation of proteolysis. The differential effects of Doa4 inactivation on various substrates (Fig. 5) would thus reflect the relative abilities of these substrates to compete for limiting amounts of active protease. Incomplete proteolytic inhibition could also be due to an additional deubiquitinating enzyme(s) that partially compensates for loss of Doa4. Growth

of mutants deleted for multiple deubiquitinating enzymes, including *doa4 ubp1 yuh1*, *doa4 ubp1 ubp2 ubp3* and all possible double and triple mutants involving *doa4* and *ubp1-3*, is identical to that of an otherwise isogenic *doa4* strain (unpublished data), suggesting that this activity, if it exists, is not due to Ubp1-3 or Yuh1. Several deubiquitinating activities have been described in rabbit reticulocytes that appear to function with the 26S proteasome⁶. It will be of interest to determine whether any of these activities is due to the rabbit homologue of Doa4.

The TRE210/ORF1 product of the human *tre-2* gene is tumorigenic when expressed at high levels; tumorigenicity is suppressed if the TRE210/ORF1 reading frame is extended through the end of TRE213/ORF2 (ref. 20). The corresponding deubiquitination assays (Fig. 3) show that the TRE210/ORF1 oncoprotein is devoid of enzyme activity but that activity is regained in the extended TRE210/ORF1-TRE213/ORF2 polypeptide. Our data suggest a straightforward explanation of the *tre-2* tumorigenicity results, namely, that the inactive TRE210/ORF1 protein represents a truncated form of *tre-2* that interferes with the normal cellular *tre-2* product(s) in a dominant negative fashion, whereas restoration of deubiquitinating activity in the extended protein relieves this interference. Hence the human *tre-2* gene may function, at least under certain conditions, as a tumour suppressor. Several other mammalian growth regulators, for example, p53 and *erbA*, are subject to both recessive and dominant negative mutations²⁸. That the Doa4/*tre-2* class of deubiquitinating enzymes can suffer dominant negative mutations is indicated by our analysis of the *doa4*^{Ser571} allele in yeast (Fig. 5e,f). When expressed at high levels, *doa4*^{Ser571} strongly inhibits ubiquitin-dependent proteolysis and causes a pleiotropic mutant phenotype. Based on its similarity in sequence and enzymatic properties to Doa4 (Figs 2 and 3), the *tre-2* enzyme(s) may also be required for efficient ubiquitin-dependent proteolysis *in vivo*. In this view, the misregulation of growth in cells expressing oncogenic derivatives of *tre-2* would be attributed to a failure to degrade specific short-lived regulatory proteins that are involved in cell proliferation. Many proto-oncogene products, for instance, are short-lived and may be degraded by the ubiquitin system; it is known that increased levels of such proteins can be tumorigenic^{2,29,30}.

Note added in proof: It has recently been shown that the *tre-2* oncogene is derived from a family of genes on chromosome 17, now designated the TRE17 genes⁴⁶. □

Received 22 June; accepted 8 October 1993.

1. Glotzer, M., Murray, A. W. & Kirschner, M. W. *Nature* **349**, 132-138 (1991).
2. Okazaki, K., Nishizawa, M., Furuno, N., Yasuda, H. & Sagata, N. *EMBO J.* **11**, 2447-2456 (1992).
3. Hochstrasser, M., Ellison, M. J., Chau, V. & Varshavsky, A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4606-4610 (1991).
4. Chen, P., Johnson, P., Sommer, T., Jentsch, S. & Hochstrasser, M. *Cell* **74**, 357-369 (1993).
5. Finley, D. & Chau, V. *Rev. Cell Biol.* **7**, 25-69 (1991).
6. Hershko, A. & Ciechanover, A. *Rev. Biochem.* **61**, 761-807 (1992).
7. Hochstrasser, M. *Curr. Opin. Cell Biol.* **4**, 1024-1031 (1992).
8. Jentsch, S. A. *Rev. Genet.* **26**, 177-205 (1992).
9. Varshavsky, A. *Cell* **69**, 725-735 (1992).
10. Hochstrasser, M. & Varshavsky, A. *Cell* **61**, 697-708 (1990).
11. Rose, I. A. in *Ubiquitin* (ed. Rechsteiner, M.) 135-155 (Plenum, New York, 1988).
12. Hadari, T., Warme, J. V. B., Rose, I. A. & Hershko, A. *J. biol. Chem.* **267**, 719-727 (1992).
13. Murakami, Y. et al. *Nature* **360**, 597-599 (1992).
14. Rechsteiner, M. *J. biol. Chem.* **268**, 6065-6068 (1993).
15. Mayer, A. N. & Wilkinson, K. D. *Biochemistry* **28**, 166-172 (1989).
16. Jonnalagadda, S. et al. *J. biol. Chem.* **264**, 10637-10642 (1989).
17. Baker, R. T., Tobias, J. W. & Varshavsky, A. *J. biol. Chem.* **267**, 23364-23375 (1992).
18. Miller, H. I. et al. *J. 598-709* (1989).
19. Tobias, J. W. & Varshavsky, A. *J. biol. Chem.* **266**, 12021-12028 (1991).
20. Nakamura, T. et al. *Oncogene* **7**, 733-741 (1992).
21. Chau, V. et al. *Science* **243**, 1576-1583 (1989).
22. Cook, W. J., Jeffrey, L. C., Carson, M., Chen, Z. & Pickart, C. M. *J. biol. Chem.* **267**, 16467-16471 (1992).
23. Ellison, M. J. & Hochstrasser, M. *J. biol. Chem.* **266**, 21150-21157 (1991).
24. Gupta, K., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. & Gray, D. A. *Oncogene* **8**, 2307-2310 (1993).
25. Fischer-Vize, J. A., Rubin, G. M. & Lehmann, R. *Development* **116**, 985-1000 (1992).
26. Özkaynak, E., Finley, D., Solomon, M. J. & Varshavsky, A. *EMBO J.* **6**, 1429-1439 (1987).
27. Orłowski, M. *Biochem.* **29**, 10289-10297 (1990).

28. Klein, G. *FASEB J.* **7**, 821-825 (1993).
29. Miller, D. A., Curran, T. & Verma, I. M. *Cell* **36**, 51-60 (1984).
30. Small, M. B., Hay, N., Schwab, M. & Bishop, J. M. *Molec. cell Biol.* **7**, 1638-1645 (1987).
31. Struhl, K. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8419-8423 (1985).
32. Singer, V. L., Wobbe, C. R. & Struhl, K. *Genes Dev.* **4**, 636-645 (1990).
33. Rose, M. D., Novick, P., Thomas, J. H., Botstein, D. & Fink, G. R. *Gene* **60**, 237-243 (1987).
34. Gietz, R. D. & Sugino, A. *Gene* **74**, 527-534 (1988).
35. Ausubel, F. M. et al. *Current Protocols in Molecular Biology* (Wiley, New York, 1989).
36. Sharp, P. M. & Li, W.-H. *Nucleic Acids Res.* **15**, 1281-1295 (1987).
37. De Virgilio, C. et al. *Eur. J. Biochem.* **212**, 315-323 (1993).
38. Dayhoff, M. O., Barker, W. C. & Hunt, T. L. *Meth. Enzym.* **91**, 534-545 (1988).
39. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215**, 403-410 (1990).
40. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. *Science* **254**, 1374-1377 (1991).
41. Gallwitz, D. & Sures, I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2546-2550 (1980).
42. Bachmair, A., Finley, D. & Varshavsky, A. *Science* **234**, 179-186 (1986).
43. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. *Gene* **100**, 119-122 (1992).
44. Dohmen, R. J., Madura, K., Bartel, B. & Varshavsky, A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7351-7355 (1991).
45. Deveraux, Q., Wells, R. & Rechsteiner, M. *J. biol. Chem.* **265**, 6323-6329 (1990).
46. Onno, M., Nakamura, T., Mariage-Samson, R., Hillova, J. & Hill, M. *DNA Cell Biol.* **12**, 107-118 (1993).

ACKNOWLEDGEMENTS. We thank R. Berenson and D. Klaman for help in the cloning and sequencing of DOA4; M. Hill for *tre-2* cDNA clones; W. Cook for purified ubiquitin dimer; R. McCarroll for help with the γ -irradiation experiments; Q. Deveraux for anti-ubiquitin antibody; S. Van Nocker for the multiubiquitin size standards; D. Gottschling and P. Johnson for comments on the manuscript; and R. Baker for strains. This work was supported by grants to M.H. from the NIH, the Cancer Research Foundation, the Louis Block Fund, the University of Chicago Cancer Research Center, and the Searle Scholars Program of the Chicago Community Trust. F.R.P. was supported by a NIH predoctoral fellowship.

An extreme scattering event in the direction of the millisecond pulsar 1937 + 21

Ismaël Cognard, Gabriel Bourgois, Jean-François Lestrade, François Biraud, Dominique Aubry, Bernard Darchy & Jean-Pierre Drouhin

Observatoire de Paris, Section Meudon 92195, Meudon, Principal Cedex, France

LARGE fluctuations in the radio emissions from the quasar 0954 + 658 (ref. 1) attest to the existence of large-scale inhomogeneities in the ionized interstellar medium. These fluctuations, termed extreme scattering events, are caused by the refractive focusing of the radio waves by discrete plasma structures. In principle, the radio emissions of pulsars should also be a sensitive probe of such phenomena²⁻⁵, which should be manifest as fluctuations in the flux density and increased delays in the timing measurements. Here we report the detection of an extreme scattering event, lasting about 15 days, in timing observations of the millisecond pulsar 1937 + 21. This event provides independent evidence for the existence of large-scale structure in the interstellar medium, and allows us to constrain the properties of the lensing structure (distance, velocity, linear dimension and electron density) more tightly than has been possible from detections of lensed extragalactic radio sources. Of more practical importance, such events could be the dominant source of timing noise in millisecond pulsars at ~1 GHz, and will need to be considered when searching for the subtle signature of gravitational waves in pulsar timing measurements.

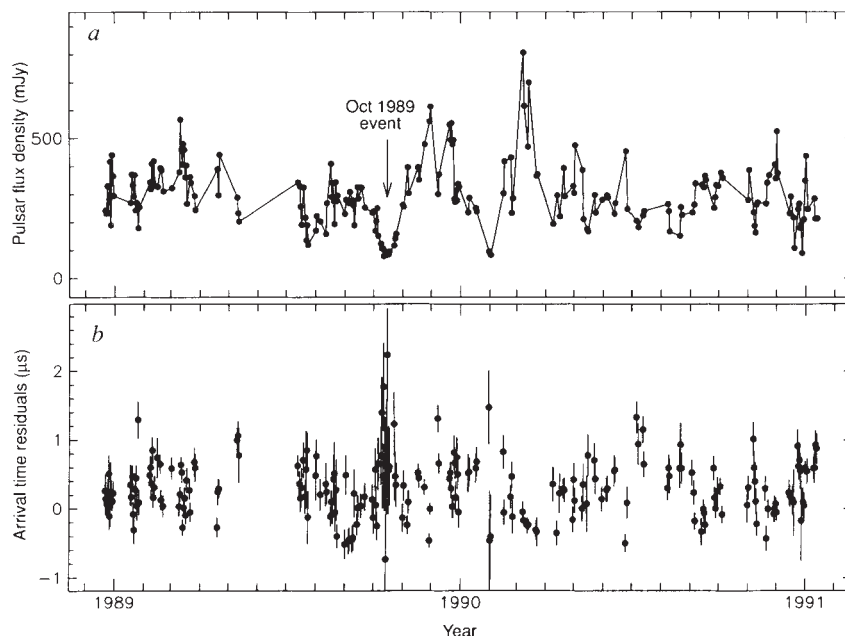
The times of arrival (TOA) of the sharp pulses of millisecond pulsars can be measured with microsecond accuracy, or better, over many years. These pulsars have been studied by several radio observatories to provide information on the cosmic gravitational background and the ionized interstellar medium. Such investigations have applications in astrometry, the stability of the atomic timescale, and in improving the ephemerides of the Solar System. The timing observations of the millisecond pulsar 1937 + 21 presented here have been conducted at the large transit radio telescope (7,000 m², equivalent to a 93-m dish) located

near Nançay (France). Since late 1988 we have conducted timing observations of PSR1937+21 ~8–12 times a month at 1,410 MHz at both senses of polarization and at 1,700 MHz since early 1992 to monitor the variations in dispersion measure (DM).

The daily timing observations of PSR1937+21 in October 1989 were conducted over a period when the flux density of this pulsar steadily decreased to reach a minimum of 80 mJy, then increased back to its typical flux density of ~350 mJy over 15 d. The analysis of the timing data has been carried out by our software ANTIOPE which includes the classical model for pulsar TOA^{6,7} with the Earth orbit geometric delay, the general relativity delay and the mean dispersion measure ($\langle DM \rangle$) delay and a least-squares procedure to adjust relevant parameters. The Earth's orbital motion is read from the Jet Propulsion Laboratory ephemerides DE202 and the timescale adopted is International Atomic Time (TAI). Diffractive and refractive perturbation terms must complement the classical TOA model to account for turbulences and possible large-scale discrete structures in the ionized interstellar medium. These perturbations have been categorized as: (1) the dispersive delay $\delta\tau_{DM}$ which is caused by DM variations around the mean value $\langle DM \rangle$; (2) the geometric refractive delay $\delta\tau_{geo}$ which is the increased path length caused by refraction in an intervening discrete structure; and (3) the barycentric refractive delay $\delta\tau_{bary}$ which is caused by change of the angle-of-arrival after refraction in the structure^{2,4,5}.

The timing observations of PSR1937+21 analysed here are from 22 December 1988 to 13 January 1991 (217 TOAs). The seven pulsar parameters solved for with ANTIOPE (period P , $\dot{P}$, coordinates, proper motion components and initial rotational phase) will be presented elsewhere, and compared to the Arecibo solution. The post-fit TOA residuals of this solution are characterized by a mean of 0.01 μ s and a standard deviation of 0.36 μ s. Fig. 1b shows all TOA residuals for the 2-yr period and it is clear that they are positively biased in October 1989 when the flux density of PSR1937+21 dropped dramatically as seen in Fig. 1a. Note also that the rapid fluctuations of the flux density for ~2 months after the event in Fig. 1a might be a related phenomenon but we shall concentrate here on the well identified part of the event. An enlargement of the event in Fig. 2 shows the striking behaviour of the TOA post-fit residuals with two cusps as high as 2 μ s. These cusps are expected to occur when the increased path length (τ_{geo}) caused by refraction reaches a maximum on each side of an axisymmetric lensing structure².

Fig. 1 Timing observations of PSR1937+21 at Nançay Radio Observatory from 22 December 1988 to 13 January 1991. a, Behaviour of the flux density, b, post-fit time of arrival (TOA) residuals after fitting the pulsar parameters characterized by a standard deviation of 0.36 μ s. The pulsar signal is de-dispersed by a swept-frequency oscillator at 80 MHz in the station intermediate-frequency link. This oscillator is a slaved voltage controlled oscillator, driven by a parabolic saw-tooth waveform, synthesized from steps of 100 ns. The station autocorrelator is used to detect the pulse in the frequency domain. The uncertainty of the flux density of the pulse peak depends on the noise set by the integration time, and is 17 mJy for the usual observation of 1 h at the same elevation all the year around at the transit radiotelescope.



This characteristic signature of $\delta\tau_{\text{geo}}$ has already been studied and simulated in conjunction with extreme scattering events (ESEs) and pulsar timing⁵ but we believe this is the first time that it has actually been detected in pulsar timing data. We ensure that the fitting procedure cannot fake this characteristic signature by performing a check as follows: we obtained fits to the two-yr data set with and without the data segment corresponding to the October 1989 event and found no statistically significant differences in the values of the seven fitted parameters. Finally, there was no significant broadening of the pulse width of PSR1937+21 (full-width half-maximum FWHM $\approx 35 \mu\text{s}$) during the event as expected, as any broadening should be of the order of the increased path delay (τ_{geo}) during the event, that is $2 \mu\text{s}$. The Crab Nebula pulsar during a period of depressed flux density has exhibited such a broadening but it was probably caused by the surrounding material of the nebula⁸ whereas an unrelated intervening structure is responsible for the PSR1937+21 event discussed here.

The combination of the flux density behaviour and the deterministic signature in the TOA post-fit residuals of PSR1937+21 form a new and unique set of data to constrain the geometric and physical properties of the ionized structure responsible for the event. It is noticeable that during the event, we were not sensitive to the barycentric refractive perturbation because the Earth barycentric position and the pulsar direction are almost orthogonal in October⁵. Consequently, the modelling of the perturbations in Fig. 2a and b is totally dominated by $\delta\tau_{\text{DM}}$ and $\delta\tau_{\text{geo}}$. These perturbations are $\delta\tau_{\text{DM}}(x) \propto \delta n_e(x)l$ and $\delta\tau_{\text{geo}}(x) = (1/2c)\theta_r(x)^2 D_s(1 - (D_s/D))$, where the refraction angle $\theta_r(x) \propto \partial\delta n_e(x)/\partial x$, D and D_s are distances to the pulsar and to the screen, respectively, $\delta n_e(x)$ is the electron density profile of the ionized structure modelled as a one-dimensional screen along x , and l is the FWHM for the gaussian electron profile $\delta n_e(x)$. We have used ray-tracing to simulate the defocusing in the far field to model the flux density behaviour. Geometric optics applies here because the Fresnel radius is $5 \times 10^{-4} \text{ AU}$ (for $\lambda = 21 \text{ cm}$ and $D_s \approx 1 \text{ kpc}$), much smaller than the astronomical unit (AU)-size expected for the structure¹. This numerical simulation corresponds partially to the analytical formula of the gain $G = (1 + D_s \partial\theta_r(x)/\partial x)^{-1}$ when $D_s \partial\theta_r(x)/\partial x > -1$ (ref. 4). We used several electron profiles $\delta n_e(x)$: lorentzian, gaussian and bubble-like with constant internal electron density. The latter (bubble-like) produces abrupt changes in the refraction angle that give a double-peaked signature much broader than the observed signature in Fig. 2b. The overall agreement of the lorentzian and gaussian profiles with the observations are not significantly different, and we adopted the gaussian profile. We have determined parameters for the intervening structure using the maximum δn_e of the electron profile, the velocity v and the distance D_s , while the linear dimension is constrained to be $l = vt_D$ with the event duration $t_D \approx 15 \text{ d}$. The recent distance determination $D = 3.6 \text{ kpc}$ for PSR1937+21 (ref. 9) has been adopted. The analytical formulae described above for the two TOA perturbations $\delta\tau_{\text{DM}}$ and $\delta\tau_{\text{geo}}$ were used to calculate the TOA excess during the event.

An extensive search in the parameter-space ($0.5 < D_s < 3.5 \text{ kpc}$, $1 < v < 50 \text{ km s}^{-1}$, $5 < \delta n_e < 4,000 \text{ cm}^{-3}$) has been carried out. For each triplet of values, residuals with respect to the observed flux density curve and the TOA excess were calculated and normalized to a flux density uncertainty of 35 mJy and a TOA uncertainty of $0.5 \mu\text{s}$ to make the χ^2 similar to the number of degrees of freedom. It was found that a two-component model for the structure, that is two superimposed gaussian electron profiles (δn_{e1} , l_1 ; δn_{e2} , l_2) with the same velocity v and distance D_s , fit satisfactorily the data in Fig. 2a and b. The minimum χ^2 for this fit was found at $\delta n_{e1} = 25^{+25}_{-20} \text{ electrons cm}^{-3}$ and $l_1 = 0.094^{+0.040}_{-0.014} \text{ AU}$ and $\delta n_{e2} = 220^{+60}_{-70} \text{ electrons cm}^{-3}$ and $l_2 = 0.050^{+0.014}_{-0.013} \text{ AU}$ at a distance $D_s = 2.8^{+0.5}_{-1.6} \text{ kpc}$ and $v = 15 \pm 2.5 \text{ km s}^{-1}$. For these values, the non-normalized χ^2 for the flux density curve is 29.4 and for the TOA excess is 24.2 which

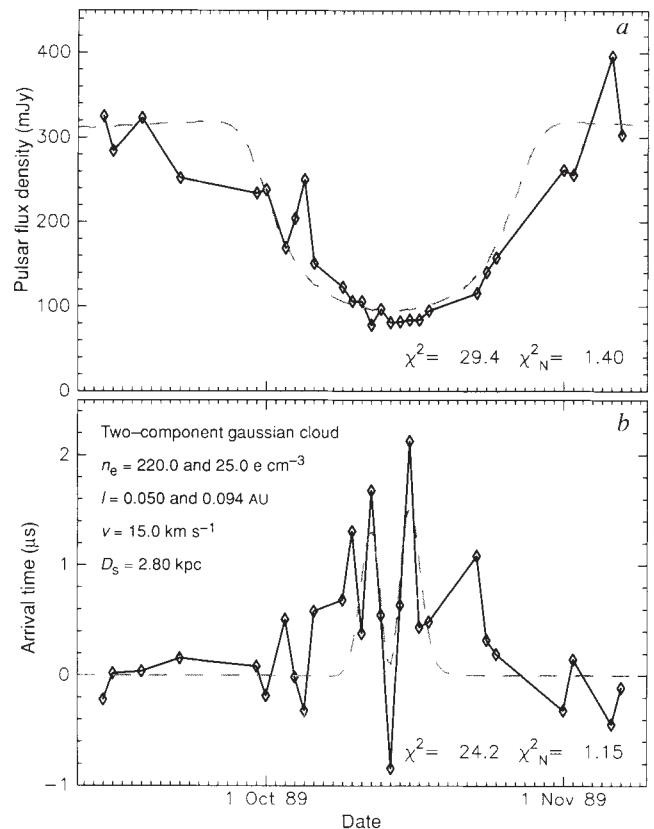


FIG. 2 Enlargement of the extreme scattering event (ESE) on PSR1937+21 in October 1989. a, The flux density behaviour and b, the post-fit TOA residuals both show the double-peaked signature expected to occur when the increased path length caused by refraction reaches a maximum on each side of an axisymmetric lensing structure². The diamonds are the data points and the dashed lines are the model of a purely refractive lens with a two-component model for the electron profile. The theoretical TOA measurement uncertainty is $0.35 \mu\text{s}$ during the event, and is $0.2 \mu\text{s}$ when the flux density is close to the averaged value of 300 mJy . χ^2_N is the χ^2 normalized by the number of degrees of freedom.

compare favourably with the number of degrees of freedom (21) for each of the two data sets. This minimum χ^2 solution was unique over the parameter-space searched. The uncertainties were calculated by varying the parameters to increment the χ^2 until the $(3 - \sigma)$ confidence level was reached. The intervening structure found in the direction of PSR1937+21 has characteristics significantly different from the ones found for the standard ESE on the radio quasar 0954+658 ($\delta n_e \approx 4,000 \text{ electrons cm}^{-3}$, $v \approx 200 \text{ km s}^{-1}$, $l \approx 7 \text{ AU}$) (ref. 1). Finally, we note that the focal distance for this model is $\sim 150 \text{ pc}$ for the shallow gaussian and 15 pc for the peaked gaussian, that is very close to the lens.

It is interesting to note that PSR1937+21 is angularly close to, and more distant than the Cygnus loop which extends from 0.5 to 2.5 kpc from the Sun and occupies the sky sector (right ascension $19 \text{ h } 20 \text{ min}$ to 22 h , declination 30° to 50°) (ref. 10). The middle-aged supernova remnants in the loop can shape the ionized interstellar medium into sheet-like structures¹¹ and it has been suggested that such a structure could be responsible for ESEs detectable in the timing of millisecond pulsars². The two-component model used in our analysis is rather *ad hoc* and models functionally closer to a sheet-like structure will be studied. The nature of the ESE observed, and the location of PSR1937+21 relative to the Cygnus loop, both support the model for ESEs produced by structures of enhanced density near the periphery of expanding superbubbles¹².

Pulsar timing observations combined with flux-density history are powerful tools to constrain the localized large-scale inhomogeneities in the ionized interstellar medium. The relatively short timescale of the ESE observed on PSR1937+21 (15 d) requires daily timing observations. By inspecting Fig. 1a, we can speculate that in early February 1990, an event on PSR1937+21 was missed because it is apparent that the flux density has dropped persistently and that the TOAs residuals are more scattered than elsewhere. As previously found in the event of October 1989, a maximum in the flux density curve is formed after the minimum, but our under-sampling makes any analysis of this event very speculative. Finally, it is possible that refractive scintillation, or superposition of miniature ESEs, might be the dominant source of noise at ~ 1 GHz in the timing of millisecond pulsars. Dense observations should indicate the deterministic signature of the larger events ($> 1 \mu\text{s}$ in TOA). $\square$

Received 15 February; accepted 22 September 1993.

1. Fiedler, R. L., Dennison, B., Johnston, K. J. & Hewish, A. *Nature* **326**, 675–678 (1987).
2. Romani, R. W., Blandford, R. D. & Cordes, J. M. *Nature* **328**, 324–326 (1987).
3. Goodman, J. J., Romani, R. W., Blandford, R. D. & Narayan, R. *Mon. Not. R. astr. Soc.* **229**, 73–102 (1987).
4. Cordes, J. M., Pridwerbetsky, A. & Lovelace, R. V. E. *Astrophys. J.* **310**, 737–767 (1986).
5. Foster, R. S. & Cordes, J. M. *Astrophys. J.* **364**, 123–135 (1991).
6. Hellings, R. W. *Astr. J.* **91**, 650–659 (1986).
7. Taylor, J. H. & Weisberg, J. M. *Astrophys. J.* **345**, 434–450 (1989).
8. Lyne, A. G. & Thorne, D. J. *Mon. Not. R. astr. Soc.* **172**, 97–108 (1975).
9. Taylor, J. H. & Cordes, J. M. *Astrophys. J.* **411**, 674–684 (1993).
10. Bochkarev, N. G. & Sitnik, T. G. *Astrophys. Space Sci.* **108**, 237–302 (1985).
11. Hester, J. J. *Astrophys. J.* **314**, 187–202 (1987).
12. Dennison, B. & Simonetti, J. H. in *Colloquium on Subarcsecond Radio Astronomy Manchester*, 20–24 July 1992 (eds Davis, R. J. & Booth, R. S.) 92–93 (Cambridge Univ. Press, 1993).

ACKNOWLEDGEMENTS. We thank G. Fréon and the Laboratoire Primaire du Temps et Fréquence (LPTF) at the Observatoire de Paris for daily corrections between the station clock at Nançay and the international UT timescale. We thank B. Dennison for comments on the manuscript. This paper is dedicated to the memory of our late co-worker, G. Bourgois.

Control of spiral-wave dynamics in active media by periodic modulation of excitability

Oliver Steinbock, Vladimir Zykov* & Stefan C. Müller

Max-Planck-Institut für molekulare Physiologie, Rheinlanddamm 201, D-44139 Dortmund 1, Germany

EXCITABLE media exhibit a wide variety of geometrically complex spatio-temporal patterns, perhaps the most striking of which are rotating spiral waves. Spiral waves have now been observed in many excitable systems, including heart muscle¹, aggregating slime-mould cells², retinas³, CO oxidation on platinum⁴ and oscillatory chemical systems such as the Belousov–Zhabotinsky (BZ) reaction^{5,6}. In the last case, the spiral cores trace out circular or hypocycloidal trajectories, depending on the specific reaction conditions^{7–9}. In addition, if the excitability of the BZ reaction is light-sensitive^{10–13}, constant illumination has been shown to influence the dynamics of spiral waves^{14,15}. Here we investigate the effect of illumination that is periodically modulated in time. We find that, for a single set of reaction conditions, the motion of the spiral cores can be forced to describe a wide range of open and closed hypocycloidal trajectories, in phase with the applied modulation frequency. Numerical simulations using a modified version of the Oregonator model^{16,17} of the BZ reaction reproduce this behaviour. We suggest that the modulation of excitability with weak external forces might be used as a means for controlling the dynamics of other excitable media.

In the light sensitive BZ reaction^{10,11}, a ruthenium-bipyridyl complex is used to promote the autocatalytic production of the reaction activator HBrO_2 . This only occurs when the complex is in its reduced, and electronically unexcited state. If the ruthenium complex is photochemically excited, it slowly catalyses the production of the inhibitor species, bromide. Thus externally applied illumination suppresses the excitability of the medium (for instance, it decreases the propagation velocity of excitation waves) and allows control of spiral wave parameters^{12–15}.

In our experiments $\text{Ru}(\text{bpy})_3^{2+}$ (4 mM) was immobilized in a silica-gel matrix¹⁸ (thickness 0.7 mm, diameter 7 cm). The reactants and their concentrations (disregarding bromination of malonic acid) were; NaBr (0.09 M), NaBrO_3 (0.19 M), malonic acid (0.17 M) and H_2SO_4 (0.35 M). The temperature was kept constant at $23 \pm 1^\circ\text{C}$. White light (halogen lamp, 150 W) illumi-

nating the entire observation area was polarized by a rotating polarization filter and applied to the active medium through a tilted glass plate. Because of the rotation of the polarization vector the intensity of the reflected light was modulated sinusoidally with time (0.49 – 1.36 mW cm^{-2}). The transmitted light was detected by CCD (charge-coupled device) camera (Hamamatsu C3077) at 490 nm, stored on a video recorder and finally digitized by an image-acquisition card.

We created a spiral wave by using a thin laser beam to break a propagating circular wave (this creates two spirals) and moving one of the open ends to the boundary of the dish¹⁹. The temporal trace of the wave tip was detected visually with a reticle in digitized images. Figure 1 shows a snapshot of the created spiral wave (constant light intensity, 0.93 mW cm^{-2}) with the overlaid trajectory of its tip. The trajectory is almost a five-lobed hypocycloid with the wave period $T_0 = 24.5 \text{ s}$ at the centre of the meandering pattern. The change in the stationary intensity level of illumination in the range from 0.49 to 1.36 mW cm^{-2} leads only to small variations of the spatio-temporal parameters of the spiral wave. For the minimum and maximum intensities applied here, we observed four- and six-lobed trajectories, respectively.

Periodic modulation of the applied light intensity within the same range causes dramatic changes in the dynamics of spiral rotation. The small variations of the trajectory shape that occur during each period of the modulation accumulate with time. As a result, the modulation forces the tip to follow trajectories that

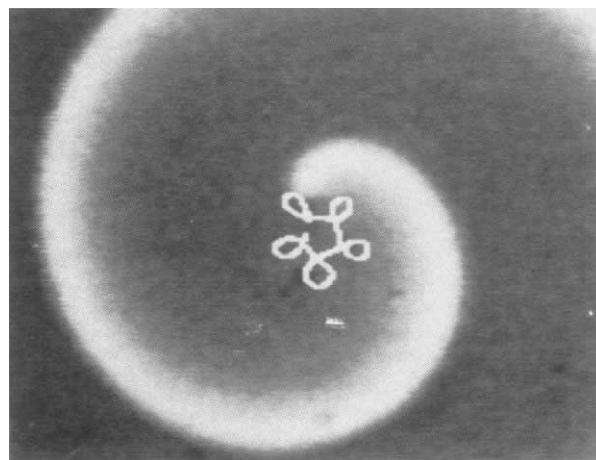


FIG. 1 Spiral wave in the ruthenium-catalysed BZ reaction at a steady level of light intensity (0.93 mW cm^{-2}). In the experiment bright bands are green, and dark bands are orange. The tip trajectory (overlaid white curve) is similar to a hypocycloid. Image size $3.8 \times 3.0 \text{ mm}$.

* Permanent address: Institute of Control Sciences, Moscow, Russia.

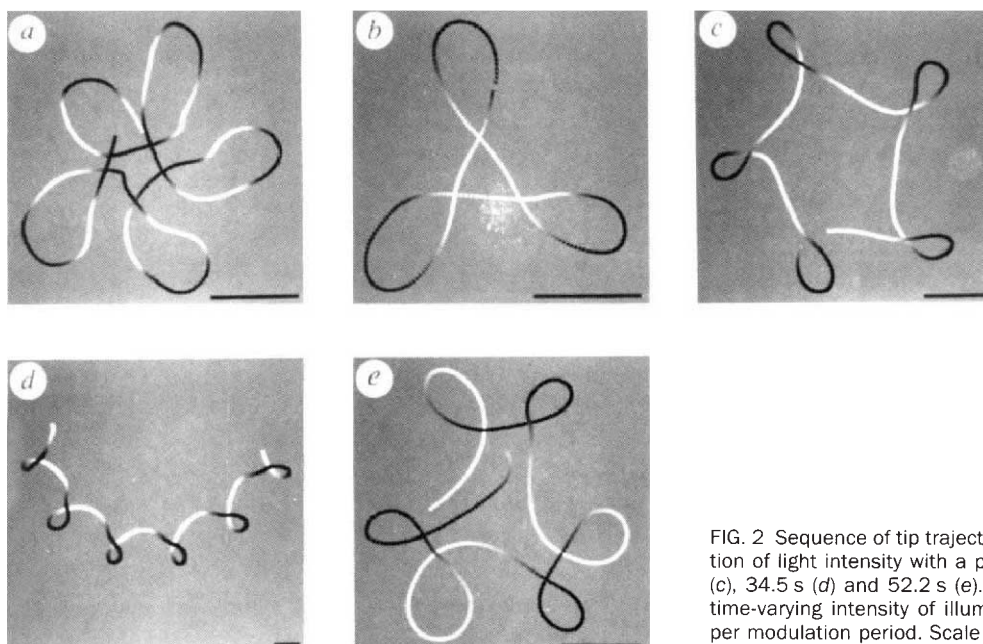


FIG. 2 Sequence of tip trajectories measured under sinusoidal modulation of light intensity with a period T_m of: 17.0 s (a), 26.2 s (b), 30.4 s (c), 34.5 s (d) and 52.2 s (e). The shading of the traces is due to the time-varying intensity of illumination and shows the number of lobes per modulation period. Scale bar, 0.2 mm.

differ significantly from those observed at constant intensities. The shape of trajectories depends strongly on the modulation period T_m (see Fig. 2). The trajectories of Fig. 2b–d are members of an entrainment band ($T_m \approx 20$ –35 s) of closed hypocycloids with one lobe corresponding to one external period. The number of lobes continuously increases from 3 to more than 12 with increasing values of T_m . Figure 2a shows a deformed five-lobed pattern with one lobe described during two modulation periods. Figure 2e depicts a surprising trajectory with alternating distances between neighbouring lobes. In this small frequency range the spiral tip describes a pair of lobes during one external modulation. For modulation periods between those of Fig. 2d and e we observed irregular motion with epicyclic segments of the trajectories. In all examples of Fig. 2 the tip motion is synchronized by the external rhythm.

We complemented the experiments by numerical simulations of the observed process of synchronization. For this purpose we used an Oregonator model^{16,17}, which was extended by a term ϕ describing the light-induced bromide production¹¹.

$$\frac{\partial u}{\partial t} = \Delta u + \frac{1}{\varepsilon} \left[u - u^2 - (fv + \phi) \frac{u - q}{u + q} \right]$$

$$\frac{\partial v}{\partial t} = u - v$$

The variables u and v describe the time-space evolution of the HBrO_2 and catalyst concentration, respectively. While the parameters $\varepsilon = 0.05$, $q = 0.002$ and $f = 2.0$ are constant, the term $\phi(t) = 0.01 + A \sin(2\pi t/T_m)$ describes the periodic modulation of bromide production, and hence excitability.

In our calculations we studied the trajectories of the spiral tip for different values of the modulation period T_m and amplitude A . The unperturbed trajectory ($A = 0$) is similar to a five-lobed hypocycloid (see Fig. 1) and the excitation period measured at the symmetry centre is $T_0 = 2.8$. Our calculations show that the motion of the spiral tip can be synchronized by external modulation. In Fig. 3 the most pronounced entrainment bands are shown (labelled 1:3, 1:2, 1:1 and 2:1), with the broadest band observed around the modulation period $T_m = T_0$. For this band one loop is completed during one modulation period. Each modulation period for the furthest right (2:1) band in Fig. 3 contains two loops. On the left side of the diagram two bands (1:3 and 1:2) are shown which correspond to five-lobed trajectories but

deformed with respect to the unperturbed one. Each loop of these meandering pictures is described during two (1:2 band) or three modulation periods (1:3 band). Within the entrainment bands the tip trajectory is sometimes rather complex but regular, in that the motion is still phase-locked. Between the bands, irregular behaviour apparently dominates. The simulation results for $A = 0.005$ are in good qualitative agreement with the experimental data given in Fig. 2. One can see the wide entrainment band ($T_m \approx T_0$) in which the lobe number increases from three up to infinity (as in Fig. 2b–d), the deformed five-lobed trajectory for $T_m \approx T_0/2$ (as in Fig. 2a) and characteristic pictures with pairs of petals for $T_m \approx 2 \times T_0$ (see Fig. 2e).

The nature of the described synchronization phenomenon is similar to entrainment in different kinds of oscillating systems in nonlinear physics²⁰ or chronobiology²¹, but cannot be reduced to the case of zero spatial dimension. The resulting family of very unusual spatio-temporal patterns, including the transition from localized to infinite motion, has no close analogies with

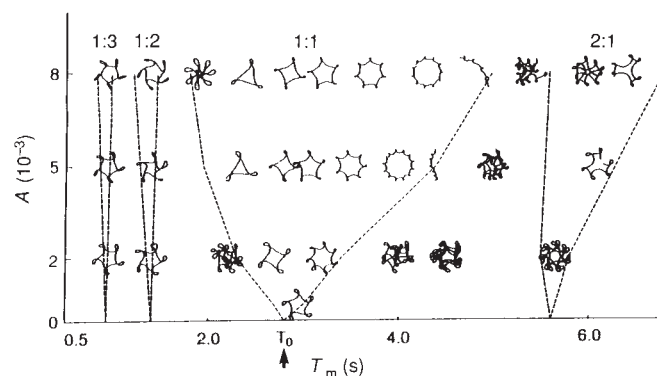


FIG. 3 Tip trajectories calculated as a response to sinusoidal modulation of bromide production. Trajectories (not to scale) are shown in the plane spanned by amplitude A and period T_m of modulation. Dashed lines indicate boundaries of entrainment bands related to different ratios $n:m$, where n is the number of lobes per m periods of the modulation. The vertical arrow on the x axis shows T_0 , the intrinsic period of the unperturbed meandering spiral.

other examples of entrainment. The observed effects also differ essentially from the resonance drift induced by periodic forcing of an active medium¹² which occurs without synchronization and leads to displacement of the spiral-wave core as a whole, without the pronounced deformations of the tip trajectory²² characteristic of the described phenomenon. This new way to deform and translate a spiral wave by controlling excitability with even weak external forces could ultimately prove relevant to our understanding of pattern formation in the early stages of morphogenesis², or to the origin and possible control of abnormalities of heart rhythm¹. □

Received 13 May; accepted 1 October 1993.

1. Davidenko, J. M., Pertsov, A. V., Salomonsz, R., Baxter, W. & Jalife, J. *Nature* **355**, 349–351 (1992).
2. Siegert, F. & Weijer, C. J. *Cell Sci.* **93**, 325–335 (1989).
3. Gorelova, N. A. & Bures, J. J. *Neurobiol.* **14**, 353–363 (1983).
4. Jakubith, S., Rotermund, H. H., Engel, W., von Oertzen, A. & Erth, G. *Phys. Rev. Lett.* **65**, 3013–3016 (1990).

5. Winfree, A. T. *Science* **175**, 634–636 (1972).
6. Field, R. J. & Burger, M. (eds) *Oscillations and Traveling Waves in Chemical Systems* (Wiley, New York, 1985).
7. Jahnke, W., Skaggs, W. E. & Winfree, A. T. *J. phys. Chem.* **93**, 740–749 (1989).
8. Plesser, Th., Müller, S. C. & Hess, B. *J. phys. Chem.* **94**, 7501–7507 (1990).
9. Skinner, G. S. & Swinney, H. L. *Physica D* **48**, 1–16 (1991).
10. Kuhnert, L. *Naturwissenschaften* **73**, 96–97 (1986).
11. Krug, H.-J., Pohlmann, L. & Kuhnert, L. *J. phys. Chem.* **94**, 4862–4866 (1990).
12. Agladze, K. I., Davydov, V. A. & Mikhailov, A. S. *JETP Lett.* **45**, 767–770 (1987).
13. Kuhnert, L., Agladze, K. I. & Krinsky, V. I. *Nature* **337**, 244–247 (1989).
14. Markus, M., Nagy-Ungvarai & Hess, B. *Science* **257**, 225–228 (1992).
15. Braune, M. & Engel, H. *Chem. Phys. Lett.* (in the press).
16. Field, R. J. & Noyes, R. M. *J. phys. Chem.* **60**, 1877–1884 (1974).
17. Jahnke, W. & Winfree, A. T. *Int. J. Bifurcation and Chaos* **1**, 445–466 (1991).
18. Yamaguchi, T., Kuhnert, L., Nagy-Ungvarai, Zs., Müller, S. C. & Hess, B. *J. phys. Chem.* **95**, 5831–5837 (1991).
19. Steinbock, O. & Müller, S. C. *Physica A* **188**, 61–67 (1992).
20. Schuster, H. G. *Deterministic Chaos* (VCH, Weinheim, 1989).
21. Hildebrandt, G., Gutenbrunner, Ch. & Moog, R. (eds) *Chronobiology and Chronomedicine* (Peter Lang, Frankfurt, 1992).
22. Davydov, V. A., Zykov, V. S. & Mikhailov, A. S. *Soviet. Phys. Usp.* **34**, 665–684 (1991).

ACKNOWLEDGEMENTS. This work was supported by the European Community (DG XII) and the Boehringer Ingelheim Fonds, Stuttgart.

Self-assembling organic nanotubes based on a cyclic peptide architecture

M. Reza Ghadiri, Juan R. Granja,
Ronald A. Milligan, Duncan E. McRee
& Nina Khazanovich

Departments of Chemistry, Molecular Biology, and Cell Biology,
The Scripps Research Institute, La Jolla, California 92037, USA

HOLLOW tubular structures of molecular dimensions may offer a variety of applications in chemistry, biochemistry and materials science. Concentric carbon nanotubes^{1,2} have attracted a great deal of attention, while the three-dimensional tubular pore structures of molecular sieves have long been exploited industrially^{3–8}. Nano-scale tubes based on organic materials have also been reported previously^{9–13}. Here we report the design, synthesis and characterization of a new class of organic nanotubes based on rationally designed cyclic polypeptides. When protonated, these compounds crystallize into tubular structures hundreds of nanometres long, with internal diameters of 7–8 Å. Support for the proposed tubular structures is provided by electron microscopy, electron diffraction, Fourier-transform infrared spectroscopy and molecular modelling. These tubes are open-ended, with uniform shape and internal diameter. We anticipate that they may have possible applications in inclusion chemistry, catalysis, molecular electronics and molecular separation technology.

The proton-triggered self-assembly process described here is a highly convergent approach in which numerous ring-shaped peptide subunits interact through an extensive network of hydrogen bonds to form nanotube structures (Fig. 1). The present design features an eight-residue cyclic peptide with the following sequence: *cyclo*[(D-Ala-Glu-D-Ala-Gln)₂]. In designing the subunit, we postulated that cyclic peptides with an even number of alternating D- and L-amino acids can adopt (or sample) a low-energy ring-shaped flat conformation in which all backbone amide functionalities lie approximately perpendicular to the plane of the structure. In this conformation, subunits can stack in an antiparallel fashion and participate in backbone–backbone intermolecular hydrogen bonding to produce a contiguous β-sheet structure. Moreover, because of the alternating D- and L-amino acid sequence, peptide side chains must necessarily lie on the outside of the ensemble thereby creating the desired hollow tubular core structure. Here, we chose to exploit the ionization state of the glutamic acid side chain functionality as the trigger

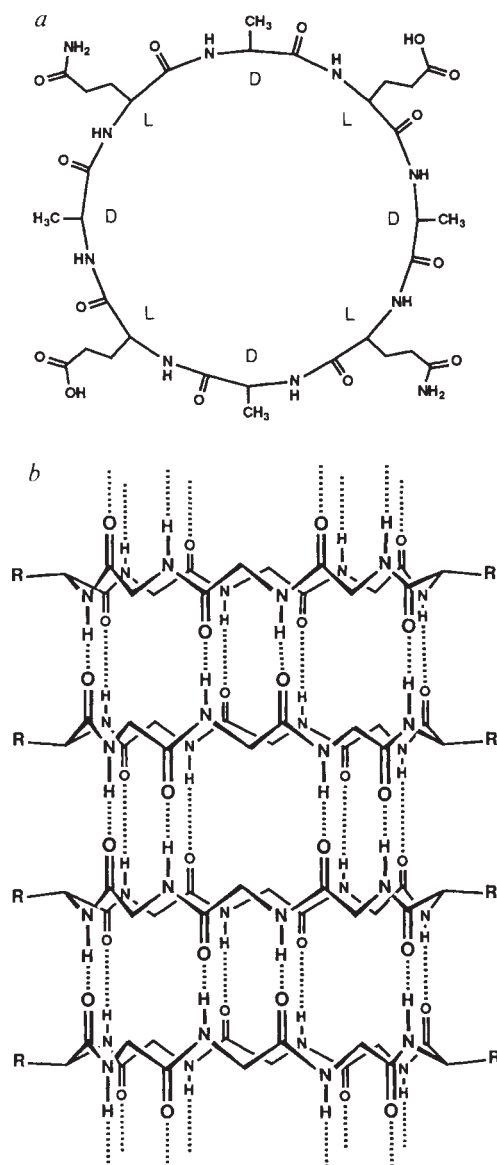
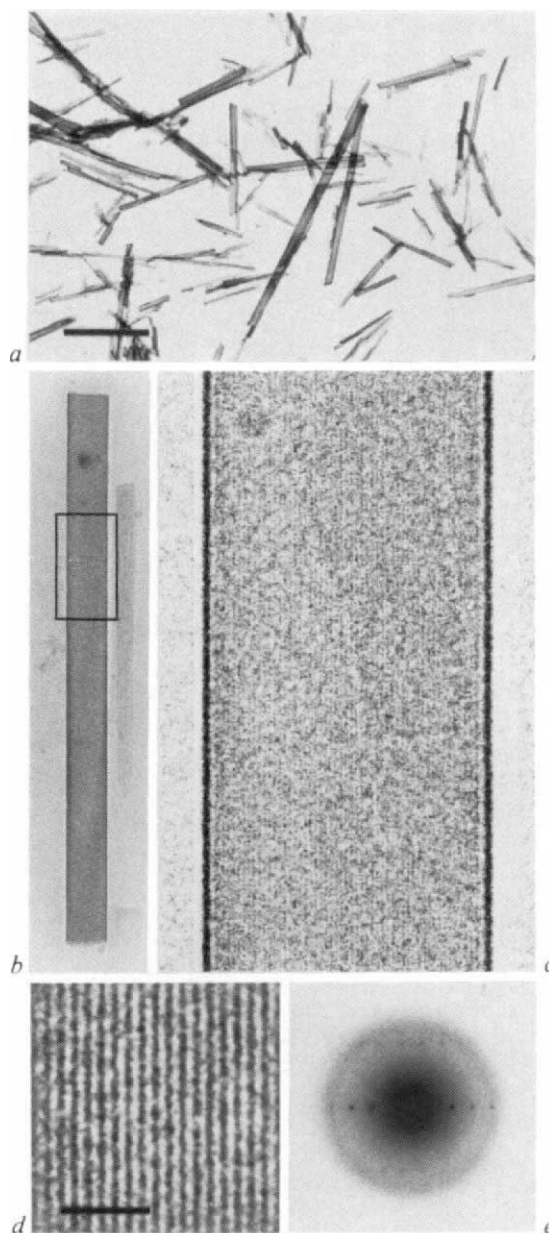


FIG. 1 a, A two-dimensional representation of the chemical structure of the peptide subunit (D or L refers to the amino acid chirality). b, Peptide subunits are shown in a self-assembled tubular configuration emphasizing the antiparallel stacking and the extensive network of intermolecular hydrogen-bonding interactions (for clarity only backbone structure is represented).

FIG. 2 Electron microscopy and electron diffraction of a nanotube particles. *a*, Low-magnification image of nanotube suspension adsorbed on carbon support film (scale bar, 1 μm). *b*, Low-dose image of a frozen hydrated single nanotube particle. The particle measures $\sim 86 \times 1180$ nm. The boxed region is enlarged in *c*, where the longitudinal striations can be seen most easily by viewing from the bottom at a glancing angle. Radiation damage and mechanical instabilities in the cold stage limit the attainable resolution in these images to ~ 10 Å. *d*, Image enhancement (see below) clearly reveals the 14.9-Å longitudinal striations. Presumably, this is due to the side-by-side packing of nanotubes in the particle (scale bar, 10 nm). *e*, Electron diffraction pattern recorded from a single nanotube particle showing orders of a 14.92-Å meridional spacing and a 4.73-Å axial spacing. Axially the pattern extends weakly to the third order (1.57 Å, data not shown) demonstrating that the nanotube particles are highly ordered and crystalline.

METHODS. The peptide was synthesized by the solid phase method²⁴ and characterized by ^1H -NMR spectroscopy, elemental analysis and ion-spray mass spectrometry. Although a variety of conditions may be used in the self-assembly of organic nanotubes, the following procedure has provided the most consistent results. A suspension (~ 25 mg ml^{-1}) of peptide subunit was clarified by the addition of 2.5 equivalents of NaOH. The resulting peptide solution was centrifuged to remove traces of solid matter and then acidified by the addition of one-third its volume of 1% trifluoroacetic acid in acetonitrile. Nanotube particles gradually form as a white suspension over a period of hours. Nanotubes may then be collected by centrifugation and washed repeatedly with distilled water to remove excess acids and salts. For electron microscopy and diffraction studies, a suspension of nanotube particles was exposed to ultrasound briefly and small drops applied to glow-discharged carbon support films on electron microscope grids. Excess liquid was removed by blotting, and the grids frozen in liquid ethane slush^{25,26}. Grids were mounted in a Gatan cold stage and examined in a Philips CM12 electron microscope operating at 120 kV. The specimen temperature was -175°C during examination and imaging. Images were recorded at $\times 35,000$ using strict low-dose conditions at various defocus levels. For image analysis, micrographs were converted to optical-density arrays using a Perkin-Elmer scanning microdensitometer with spot and step sizes equal to 2.86 Å at the specimen. Using the SUPRIM program package²⁷, a number of small areas from a single TEM image were rotationally and translationally aligned and then averaged.



mechanism for the initiation of the self-assembly process in aqueous solutions. It was expected that at alkaline pH, the large repulsive intermolecular electrostatic interactions between the negatively charged carboxylate side chains would disfavour ring stacking and at the same time promote the dissolution of the peptide subunit in aqueous media. We postulated that, on protonation of the carboxylate groups, not only would the unfavourable intermolecular electrostatic interactions vanish, but attractive side-chain/side-chain hydrogen bonding should occur. Furthermore, at acidic pH the peptide subunit is expected to be less soluble in aqueous media which was thought to contribute further to an ordered phase transition toward self-assembled nanotube particles. We therefore concluded that as intersubunit hydrogen bonding provides the principal driving force in the assembly process, and considering that only in the stacked configuration can the peptide subunits enjoy the maximum possible number of hydrogen-bonding contacts, protonation of the carboxylate groups should strongly bias the system towards self-assembly.

Controlled acidification of alkaline peptide solutions triggers the spontaneous self-assembly of peptide subunits into rod-shaped crystals (Fig. 2*a* and *b*). Transmission electron micro-

scopy (TEM) indicates that each crystal is an organized bundle of hundreds of tightly packed nanotubes. Low-dose cryomicroscopy of the crystals reveals longitudinal striations with spacing of 14.9 Å along the long axis of the crystal as expected for the centre to centre spacing for closely packed nanotubes (Fig. 2*c* and *d*). Electron diffraction patterns show 4.73-Å axial periodicity (Fig. 2*e*), suggesting that each nanotube is made up of stacked rings with intersubunit distances corresponding to an ideal antiparallel β -sheet structure^{14,15} (Fig. 1*b*). Given the thin rod shape of the crystals—on average 10–30 μm long and 100–500 nm wide—it is reasonable to assume that crystals would lie on the supporting electron microscope carbon film along the long (*a*) axis and either the *a*–*b* or *a*–*c* faces. All micrographs and diffraction patterns, however, show similar diffraction intensities and nearly identical lattices. This suggests that the *b*- and *c*-axes are identical in size within experimental error. Furthermore, given the highly symmetric nature of the octapeptide, it seems reasonable that the similarity of the views reflects the symmetry of the molecule. Electron diffraction patterns showed a unit cell having a meridional spacing of 14.92 ± 0.08 Å and an axial spacing of 4.73 ± 0.02 Å with an angle of $99.2 \pm 0.5^\circ$. The diffraction patterns did not show any symmetry other than the centre of

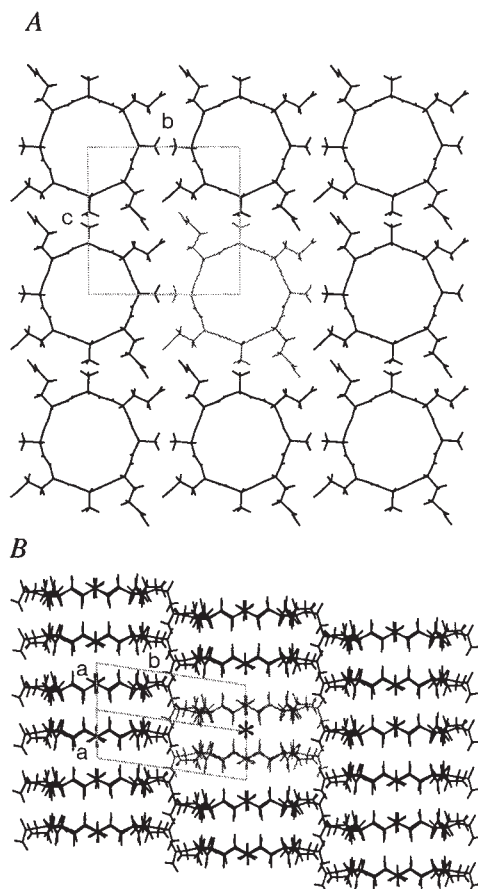


FIG. 3 Three-dimensional model of the self-assembled organic nanotubes (see text for the details of model building). View of the crystal packing along the *a*-axis (A) and the *c*-axis (B) are shown. The unit cell is indicated by the solid lines and the local 2-fold axis by the asterisk. The unit cell has dimensions $a=9.5$ Å, $b=c=15.1$ Å, and $\alpha=90^\circ$, $\beta=\gamma=99^\circ$. The tube axis is along *a*.

symmetry due to Friedel's law. The patterns extend to 1.5 Å and occasionally to 1.2 Å indicating that the crystals are highly ordered. Considering the above observations, it follows that the lattice is trigonal with $a=4.73$ Å, $b=c=15.1$ Å, $\alpha=90^\circ$ and $\beta=\gamma=99^\circ$.

Although many conformers are possible using reasonable peptide backbone ϕ , ψ angles, only one fits the box defined by the unit cell—the disk-shaped flat conformation where all side chains extend outwards in the plane of the disk. Other conformations are puckered, where the irregular presentation of side chains and backbone conformation would prevent the packing of peptide subunits to within 4.7 Å due to prohibitive van der Waals contacts. The only way to pack two peptides to within 4.7 Å is to have the backbone carbonyl functionalities hydrogen-bond intermolecularly to the nitrogen amide moieties of the opposite chain and position the side chains in the plane of the peptide backbone in order to avoid unfavourable side chain/side-chain steric interactions. Furthermore, owing to the use of alternating D- and L-amino acid residues, the most favourable intermolecular hydrogen bonds form when the disks are stacked on top of one another in an antiparallel fashion giving rise to local 2-fold symmetry (Fig. 3). The model built in this way, using the program XtalView¹⁶, has unit cells that might at first glance seem twice as large ($9.5 \times 15.1 \times 15.1$), which would require a spacing of 9.4 Å in the diffraction pattern. But when diffraction patterns are calculated, every other level of *h* cancels out due to the pseudo-symmetry of the dimer. That is, the ring flipped over is

nearly equivalent to the unflipped ring (Figs 1 and 3). Several other packing arrangements were tried but none were successful in forming a three-dimensional lattice without spatial overlaps and simultaneously fitting the spacing indicated by the electron diffraction patterns. Cells where the local 2-fold symmetry axis of the dimer is crystallographic were also considered but then the lattice could either not be modelled or the diffraction pattern had to have a mirror plane which is not observed. One other possibility is to consider that the observed diffraction pattern shows a diagonal of the lattice instead of a principle axis. This

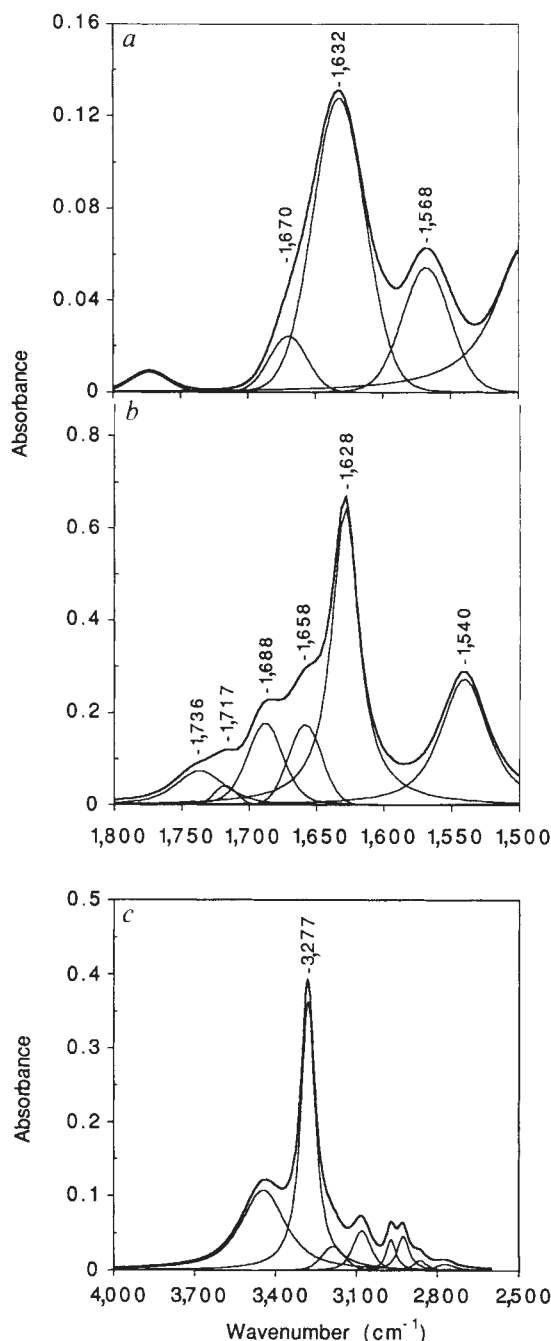


FIG. 4 Infrared spectra (at 8-cm^{-1} resolution) of amide-I region of *a*, monomeric peptide subunit in D_2O (4×10^{-3} M, $\text{pD}=10$, examined by attenuated total reflectance and *b*, nanotube particles (KBr pellet). *c*, N-H stretch region of the FT-IR spectrum of nanotube particles (KBr pellet). Component peaks were obtained by deconvolution of the original spectrum with single component, mixed lorentzian and gaussian functions using an iterative, linear least-squares algorithm (FIT, Galactic Industries).

can be easily dismissed because it requires a substantially smaller unit cell. It should be pointed out that the volume of the unit cell in our proposed model is just large enough to contain the octapeptide dimer and anything smaller can not physically hold the entire mass of the octapeptide. Therefore, the observed electron diffraction pattern must be due to the principle axis. Finally the crystalline model was checked for bad contacts using the program XPLO¹⁷ (molecular dynamics calculations on the three-dimensional crystal lattice). No bad contacts were found and the overall energy of the model was low. Most importantly, when the three-dimensional model was used to calculate structure factors, the patterns produced compared very favourably with the electron diffraction patterns thus strongly supporting the proposed model.

Involvement of intermolecular hydrogen bonding networks in tube assembly is also supported by Fourier-transform infrared (FT-IR) spectroscopic analysis¹⁸ (Fig. 4). In alkaline solutions, the monomeric peptide subunit displays an amide-I band at 1,632 cm⁻¹ consistent with the cyclic structure of the backbone. Moreover, the bands at 1,568 and 1,670 cm⁻¹ are typical solvent-exposed carbonyl stretching frequencies for glutamate and glutamine side-chain functionalities, respectively. On self-assembly, however, organic nanotubes display characteristic features of a hydrogen-bonded β -sheet structure. Not only is the appearance of amide-I bands at 1,628 cm⁻¹, and 1,688 cm⁻¹, and the amide-II band at 1,540 cm⁻¹ consistent with the expected backbone structure, but the observed N-H stretching frequency at 3,277 cm⁻¹ also strongly supports formation of a tight network of backbone-backbone hydrogen bonding. Given the tight backbone-backbone inter-subunit interactions, as determined by the above analysis of the electron diffraction patterns, it is possible to use Krimm's correlation to estimate the intermolecular hydrogen-bonding distance from the observed frequency of the N-H stretching vibration¹⁸. The observed frequency of N-H stretching mode approximately correlates to an average inter-subunit N-O distance of 2.85 Å, or an inter-subunit distance of 4.71 Å. This value is consistent with the value of 4.73 Å obtained from electron diffraction patterns. It should be pointed out that the IR spectrum of the nanotubes also closely resemble the ones reported for gramicidin A—a naturally occurring linear peptide composed of hydrophobic amino acids with alternating D- and L-configuration which is known to form dimeric β -helical transmembrane ion channel structures^{19,20}. Gramicidin A ($\uparrow\downarrow\beta^{6,4}$, that is, an antiparallel β -helix with a 6.4 residue rise per turn) has amide-I bands at 1,630, 1,685 cm⁻¹, an amide-II band at 1,539 cm⁻¹, and an N-H stretching frequency at 3,285 cm⁻¹ (ref. 21).

The general strategy described here should allow the design and synthesis of a wide range of tubular structures with specified internal diameters and surface characteristics, and having wide range of applications^{22,23}. For example, such tubular materials could be designed to mimic biological channels and pore structures, used to study physical and chemical properties of confined molecules, control growth and properties of inorganic and metallic clusters, or design novel optical and electronic devices. □

14. Salemme, F. R. *Prog. Biophys. molec. Biol.* **42**, 95–133 (1983).
15. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. *J. molec. Biol.* **226**, 1143–1159 (1992).
16. McRee, D. E. *J. Molec. Graphics* **10**, 44–46 (1992).
17. Brunger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
18. Krimm, S. & Bandekar, J. in *Advances in Protein Chemistry* (eds Anfinsen, C. B., Edsall, J. T. & Richards, F. M.) 181–364 (Academic, Orlando, 1986).
19. Wallace, B. A. & Ravikumar, K. *Science* **241**, 182–187 (1988).
20. Lango, D. A. *Science* **241**, 188–191 (1988).
21. Naik, V. M. & Krimm, S. *Biophys. J.* **49**, 1147–1154 (1986).
22. Whitesides, G. M., Mathias, J. P. & Seto, C. T. *Science* **254**, 1312–1319 (1991).
23. Ozin, G. A. *Adv. Mater.* **4**, 612–649 (1992).
24. Rovero, P., Quartara, L. & Fabbri, G. *Tetrahedron Lett.* **32**, 2639–2642 (1991).
25. Adrian, M., Dubochet, J., Lepault, J. & McDowell, A. W. *Nature* **308**, 32–36 (1984).
26. Milligan, R. A., Brisson, A. & Unwin, P. N. T. *Ultramicroscopy* **13**, 1–10 (1984).
27. Schroeter, J. P. et al. *J. Struct. Biology* **109**, 235–247 (1992).

ACKNOWLEDGEMENTS. We thank C. Choi for peptide synthesis, L. Tadesse, D. Hecht and C. Shevlin for assistance in FT-IR spectroscopy, G. Siuzdak for mass spectroscopic analyses, and M. Pique, A. Olson and M. Case for graphics presentations. This work was supported by the Arnold and Mabel Beckman Foundation. N.K. thanks NSERC and the Alberta Heritage Scholarship Fund for predoctoral fellowships. J.R.G. acknowledges the fellowship from the Consellería de Educación y Ordenación Universitaria (Xunta de Galicia, Spain). M.R.G. is a Searle scholar (1991–94), Beckman Young investigator (1991–1993), and Alfred P. Sloan research fellow (1993–1995).

The contribution of explosive volcanism to global atmospheric sulphur dioxide concentrations

G. J. S. Bluth*, C. C. Schnetzler†, A. J. Krueger‡
& L. S. Walter‡

* Universities Space Research Association, Code 921, NASA/GSFC, Greenbelt, Maryland 20771, USA

† Geography Department, University of Maryland, College Park, Maryland 20742, USA

‡ Earth Sciences Directorate, NASA/GSFC, Greenbelt, Maryland 20771, USA

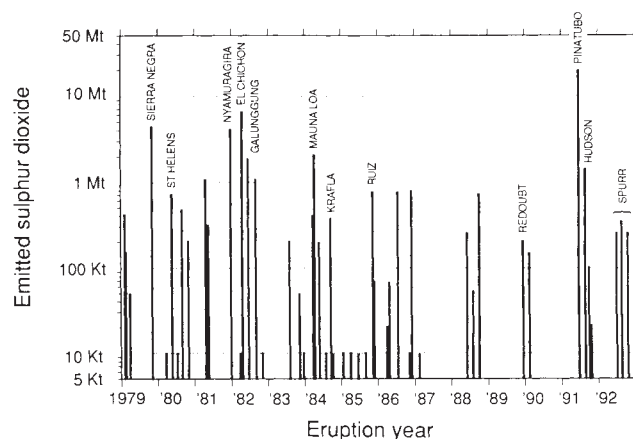
SULPHUR dioxide from volcanic eruptions may have a significant effect on the Earth's climate and atmospheric chemistry, and it is therefore important to quantify outgassing rates for all types of volcanic activity. Non-explosive volcanoes (for example, Mount Etna) outgas at relatively constant rates, providing an annual flux of about 9 million tons (Mt) SO₂ (ref. 1). By contrast, the outgassing from volcanoes prone to explosive eruptions (such as Mount Pinatubo) is sporadic and much more difficult to quantify. The total annual volcanic SO₂ flux is therefore poorly constrained, with ground-based estimates^{1–8} ranging from 1.5 to 50 Mt—up to one-quarter of the estimated current anthropogenic contribution. The Total Ozone Mapping Spectrometer aboard the NASA satellite Nimbus 7 recorded SO₂ emissions from explosive eruptions from November 1978 to May 1993. We use these data to show that the annual flux from explosive volcanism is of the order of 4 Mt SO₂, less than half of the non-explosive output. Thus it seems that the total volcanic emission of SO₂ to the Earth's atmosphere is about 13 Mt yr⁻¹, which is only 5–10% of the current anthropogenic flux.

The recent eruption of Mount Pinatubo has rejuvenated the question of volcanic effects on global climate. The climatic impact of volcanic activity is potentially large because, in addition to the relatively consistent degassing of non-explosive volcanoes into the Earth's atmosphere, the infrequent bursts of explosive volcanoes such as Pinatubo have the ability to inject sulphur dioxide directly into the stratosphere. However, constraints on the SO₂ production from global volcanism have been elusive. Previous estimates of the annual sulphur dioxide flux due to volcanic activity, using ground-based methods, vary by more than a factor of 30 (Table 1)^{1–8}. These estimates range from insignificant amounts to >25% of the present annual anthropogenic flux of ~190 million tons (Mt)⁹. Previous efforts to partition global SO₂ production into explosive and non-

Received 1 June; accepted 11 October 1993.

1. Iijima, S. *Nature* **354**, 56–58 (1991).
2. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
3. Meier, W. M. & Olson, D. H. *Atlas of Zeolite Structure Types* 2nd edn (Butterworths, London, 1988).
4. Dessau, R. M., Schlenker, J. L. & Higgins, J. B. *Zeolites* **10**, 522–524 (1990).
5. Davis, M. E., Saldarriaga, C., Montes, C., Garces, J. & Crowder, C. *Nature* **331**, 698–699 (1988).
6. Estermann, M., McCusker, L. B., Baerlocher, C., Merrouche, A. & Kessler, H. *Nature* **352**, 320–323 (1991).
7. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
8. Monnier, A. et al. *Science* **261**, 1299–1303 (1993).
9. Lotz, B., Colonna-Cesari, F., Heitz, F. & Spach, G. *J. molec. Biol.* **106**, 915–942 (1976).
10. Georger, J. H. et al. *J. Am. chem. Soc.* **109**, 6169–6175 (1987).
11. Brumlik, C. J. & Martin, C. R. *J. Am. chem. Soc.* **113**, 3174–3175 (1991).
12. Fuhrhop, J.-H., Spiroski, D. & Boettcher, C. *J. Am. chem. Soc.* **115**, 1600–1601 (1993).
13. Harada, A., Li, J. & Kamachi, M. *Nature* **364**, 516–518 (1993).

FIG. 1 Inventory of TOMS SO_2 data from explosive volcanic eruptions, 1979–1993. Most of the larger or well-known eruptions are labelled. Multiple eruptions from a single volcanic episode are combined, except for the three eruptions of Mt Spurr in 1992. The higher frequency of small eruptions observed before 1986 resulted from detailed searches through the TOMS database for all known eruptions using the Smithsonian Global Volcanism Network (GVN) database¹².



explosive activity suggest vastly different contributions: from SO_2 per year ratios of 1:14 (ref. 7) to 1:1 (ref. 1), caused primarily by their disparate evaluations of explosive outgassing. Thus, in order to determine the global volcanic emissions of sulphur dioxide, there is considerable need to constrain the production by explosive volcanoes.

Given the worldwide distribution of volcanoes in often inaccessible regions, and their unpredictable and hazardous natures, satellite observation can be an invaluable tool in assessing potential climatic effects of global volcanism. Volcanic SO_2 clouds since 1979 have been detected and tracked using data collected by the Total Ozone Mapping Spectrometer (TOMS) instrument. Total column sulphur dioxide in the Earth's atmosphere is calculated from the decrease in reflected ultraviolet sunlight due to absorption by this gas¹⁰, with an estimated error of $\pm 30\%$ ¹¹. The

TOMS, with an average ground resolution of 66 km, is the only instrument capable of measuring the full extent of SO_2 emissions from large, explosive eruptions. However, smaller events and diffuse, non-explosive emissions are often below the level of TOMS sensitivity. Another limitation exists in assessing high latitude volcanic activity in the winter (for example, eruptions in Kamchatka, Alaska and Iceland) because of decreased daylight hours, and longer path lengths due to low sun angles reduce the precision of the SO_2 algorithm; however during the past 14 years most explosive eruptions have occurred at low and mid-latitudes¹².

From 1979 through 1992, the TOMS instrument detected 55 different volcanic eruptions out of more than 350 examined, with SO_2 ranging from 10 kilotons (the approximate detection limit) to 20 Mt emitted by the recent eruption of Mount Pinatubo¹³ (Fig. 1). Some of these 55 eruptions represent consolidation of multiple events from the same volcano; for instance, 24 individual SO_2 clouds were detected from the eruption of Mount Galunggung over a six-month period in 1982 but are counted as a single volcanic event (G. J. S. B. *et al.*, manuscript submitted).

The simplest estimate of annual SO_2 output by explosive volcanism from the TOMS data is obtained from the sum of SO_2 from the 55 detected eruptions, 52 Mt, produced over 14 years of observation—an average of 4 Mt yr^{-1} . It can be seen from Fig. 1, however, that the distribution of this 'average output' is rather sporadic. The dataset is sensitive to the SO_2 emissions of the few largest eruptions. For example, one could argue that the TOMS database is simply dominated by two eruptions, El Chichón and Mount Pinatubo; removing these would halve the calculated annual SO_2 flux. Is the occurrence of these two eruptions in such a short timespan anomalous?

We may answer this question by linking the TOMS SO_2 data to a longer-term dataset, the Smithsonian Global Volcanism Network (GVN) compilation of eruption explosivity and frequency over the past 200 years. The Volcanic Explosivity Index (VEI)¹⁴ categorizes volcanic eruptions primarily from the height of the eruption column and amount of emitted tephra. Based on this scale, the TOMS instrument detected every highly explosive eruption in the past 14 years (that is, the ten eruptions of $\text{VEI} \geq 4$). Below VEI 4, the TOMS is less effective: its detection rate is $\sim 33\%$ for VEI 3 eruptions and $\sim 10\%$ for VEI 2 eruptions, using the GVN database¹² from 1979 to 1985 as the reported ground truth. The GVN data are incomplete below VEI 2, and thus we cannot directly evaluate TOMS SO_2 detection rates for VEI 0–1.

Among the TOMS-observed eruptions, there is a poor correlation between SO_2 output and explosivity of individual eruptions (Fig. 2). For example, the SO_2 emitted by El Chichón exceeded that of Mt St Helens by an order of magnitude, yet both were VEI 5 eruptions. Among the VEI 4 eruptions, TOMS SO_2 ranges over two orders of magnitude. Clearly, the SO_2 release of a given volcanic eruption cannot be inferred solely on the basis of VEI.

TABLE 1 Global sulphur dioxide flux from volcanoes (Mt yr^{-1})

Total annual	Explosive	Non-explosive	Method	Ref.
1.5			Calculated using a fixed percentage of SO_2 in magma, and total extruded material over the past 400 yr.	2
4			Same basic method as in ref. 2, but using (undefined) annual average volcanism.	3
10			Extrapolation from selected Central American volcanoes; submarine and 'very active' vents not included.	4
7.8			Using ref. 2 method but with a higher % SO_2 in the magma.	5
50			Extrapolated mainly from selected 1971–1980 volcanic activity.	6
15.2	1	14.2	Extrapolated from selected 1961–1979 volcanism to estimate several types of eruptive activity.	7
18.7	9.8	8.9	Surveyed non-explosive activity from 1981–82; extrapolated explosive activity using VEI/ SO_2 outgassing of selected eruptions.	1
50			Calculated global annual volcanic Polonium output, and used an average measured SO_2/Po ratio for several volcanoes.	8

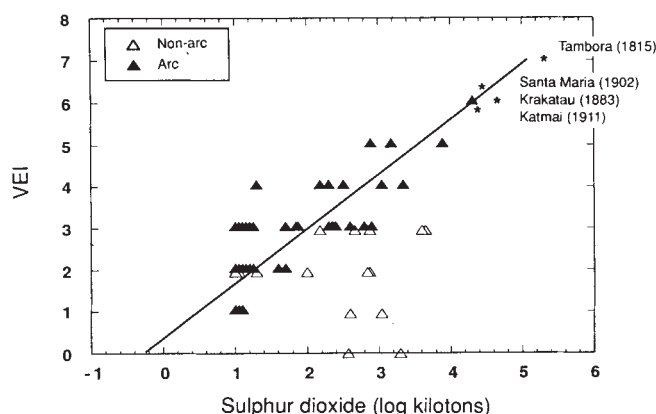


FIG. 2 SO_2 outgassing by explosive eruptions. Data are separated by tectonic regimes to show relative contributions of volcanic activity 1979–1993 as a function of explosivity. The TOMS SO_2 /VEI regression line for VEI 2–6 ($r^2=0.999$) is determined from the average SO_2 for all VEI 4–6 eruptions, whereas VEI 2 and 3 averages are calculated using the respective TOMS detection rates (see text). Estimates of SO_2 emissions (asterisks) from several historical, highly explosive eruptions are added for comparison^{16–18}.

A general relationship between average sulphur dioxide emission and VEI has been observed, however¹. Likewise in Fig. 2, a TOMS SO_2 /VEI relationship is determined from average SO_2 amounts for each VEI level (VEI 2 and 3 averages reflect the TOMS' detection rates of these eruptions). We then used the relationship of eruption frequency to VEI over the past 200 years¹⁵ to determine the annual eruptive activity for each VEI category. Multiplying the two (average SO_2 and number of eruptions from VEI 0–7) produced a total SO_2 flux from explosive volcanism of ~ 4 Mt, in agreement with the simple average of the total TOMS SO_2 database. This suggests that the dominance of the total by a few large eruptions is not unique to the past 14 years.

Differences in volcanism between non-arc (hot spot and rift zones) and arc (subduction zones) tectonic regimes are also important. Non-arc volcanoes typically erupt more frequently than arc volcanoes, but are less numerous worldwide; arc eruptions are by far the most violent¹². The TOMS data in Fig. 2 show, for arc eruptions, a general increase in outgassed SO_2 with explosivity but non-arc activity is less dependent on VEI. Non-arc volcanoes typically degas more sulphur-rich magmas and emit relatively more sulphur dioxide for a given VEI. But when considering the two regimes on a global basis, we find a 2:1 ratio of arc to non-arc SO_2 outgassing—a ratio dominated by two arc eruptions, El Chichón and Pinatubo.

The TOMS instrument gives no direct information on cloud altitude; however, we may infer a stratospheric component of the global emission of SO_2 , based on the associated explosivity of the eruptions. At VEI 3 and below, most plumes should, by definition¹⁴, remain below the tropopause. Above VEI 3, eruption clouds should reach the stratosphere: in the TOMS database, 65% of the total 52 Mt SO_2 was emitted by eruptions of VEI 4 and above (Table 2).

We can use the TOMS instrument's ability to quantify SO_2 emitted by large, explosive eruptions to evaluate previous estimates of explosive volcanism (Table 1). Our study suggests that,

even disregarding the influences on our dataset from El Chichón and Pinatubo, 1 Mt SO_2 per yr (ref. 7) is low; likewise, 9.8 Mt yr^{-1} (ref. 1) is too high. The latter determination was also calculated by relating SO_2 emissions (using data from a variety of techniques) to VEI and eruption frequency using the GVN database. We believe that the discrepancy between our degassing rates occurs because the earlier study used only SO_2 data from a small number of selected eruptions, as opposed to our inclusion of all reported eruptions during the 14-year period of TOMS operation.

The TOMS data can also help put important constraints on the SO_2 flux from total global volcanism. Based on the TOMS data, explosive eruptions produce an average of 4 Mt SO_2 per year (produced from a few large, infrequent events). The other component of the total volcanic sulphur budget, non-explosive volcanoes, degas at relatively consistent rates; the most recent evaluation of production from degassing but non-erupting and low-level, continuously erupting volcanoes (both of which cannot be evaluated from TOMS surveillance) is ~ 9 Mt SO_2 yr^{-1} (ref. 1). Combining explosive and non-explosive volcanic outgassing, the present-day volcanic contribution of SO_2 to the Earth's atmosphere can be approximated by an annual flux of 13 Mt SO_2 . Previous estimates of total SO_2 emissions from global volcanism <10 Mt yr^{-1} (refs 2, 3, 5) are probably too low. Similarly, determinations of >20 Mt SO_2 yr^{-1} (refs 6, 8) appear to overestimate global volcanism. Thus, using satellite data in conjunction with previous ground-based work, we estimate that the total volcanic contribution of SO_2 to the Earth's atmosphere totals only ~ 5 – 10% of the annual anthropogenic flux. □

Received 23 June; accepted 7 October 1993.

1. Stoiber, R. E., Williams, S. N. & Huebert, B. J. *Volcan. geotherm. Res.* **33**, 1–8 (1987).
2. Kellogg, W. W., Cadle, R. D., Allen, E. R., Lazrus, A. L. & Martell, E. A. *Science* **175**, 587–596 (1972).
3. Friend, J. P. in *Chemistry of the Lower Atmosphere* (ed. I. Rasool) 177–201 (Plenum, New York, 1973).
4. Stoiber, R. E. & Jepsen, A. *Science* **182**, 577–578 (1973).
5. Cadle, R. D. *J. geophys. Res.* **80**, 1650–1652 (1975).
6. le Guern, F. *Bull. volcan.* **45**, 3, 197–202 (1982).
7. Berresheim, H. & Jaeschke, W. *J. geophys. Res.* **88**, 3732–3740 (1983).
8. Lambert, G., Le Cloarec, M.-F. & Pennisi, M. *Geochim. cosmochim. Acta* **52**, 39–42 (1988).
9. Möller, D. *Atmos. Envir.* **18**, 19–27 (1984).
10. Krueger, A. J. *Science* **220**, 1377–1379 (1983).
11. Krueger, A. J., Walter, L. S., Schnetzler, C. C. & Doiron, S. D. *J. volcan. geotherm. Res.* **41**, 7–15 (1990).
12. Smithsonian Institution/SEAN *Global Volcanism 1975–1985* (Prentice-Hall, Englewood Cliffs, NJ, and Am. Geophys. Un., Washington DC, 1989).
13. Bluth, G. J. S., Doiron, S. D., Schnetzler, C. C., Krueger, A. J. & Walter, L. S. *Geophys. Res. Lett.* **19**, 151–154 (1992).
14. Newhall, C. G. & Self, S. *J. geophys. Res.* **87**, 1231–1238 (1982).
15. Simkin, T. A. *Rev. Earth planet. Sci.* **21**, 427–452 (1993).
16. Hammer, C. U., Clausen, H. B. & Dansgaard, W. *Nature* **288**, 230–235 (1980).
17. Stothers, R. B. *Science* **224**, 1191–1198 (1984).
18. Legrand, M. & Delmas, R. J. *Nature* **327**, 671–676 (1987).

ACKNOWLEDGEMENTS. We thank the Global Volcanism Program at the Smithsonian Institution for their contribution, S. Doiron and C. Scott for technical assistance in processing and analysing the TOMS data, and T. Gerlach, W. Rose and I. Sprod for reviewing the manuscript. This research was funded by a grant from the NASA Headquarters Volcano-Climates program.

TABLE 2 TOMS SO_2 inventory of explosive volcanism 1979–1993

VEI	Non-arc (N) (Mt SO_2)*	Arc (N) (Mt SO_2)*
0	2.4 (2)	(0)
1	1.5 (2)	0.03 (3)
2	1.6 (6)	0.12 (8)
3	9.9 (5)	3.4 (19)
4	(0)	3.8 (6)
5	(0)	10 (3)
6	(0)	20 (1)
7		
14 yr total (Mt)	15 (15)	37 (40)
Annual average (Mt yr^{-1})	1.1	2.6

* Totals from TOMS database; N, number of eruptions observed.

Fresh basalts from the Pacific–Antarctic Ridge extend the Pacific geochemical province

Elisa M. Ferguson & Emily M. Klein

Department of Geology, Duke University, Durham, North Carolina 27708, USA

OVER the past several decades, the examination of volcanic rocks recovered from mid-ocean ridges and ocean islands worldwide has led to the identification of large-scale geochemical provinces reflecting compositionally distinct domains in the Earth's mantle^{1,2}. The spatial distribution of these domains may reveal global patterns of upper-mantle convection and mixing. Previous studies have shown, for example, that the distinct Indian Ocean isotope province¹ has a relatively sharp eastern boundary within the Australian–Antarctic Discordance (AAD) south of Australia^{3,4} (Fig. 1). The juxtaposition of Indian Ocean basalt compositions west of this boundary and Pacific compositions to the east suggests that the AAD may overlie a zone of convergence between ocean-basin-scale upper-mantle convection regimes: if this is so, Pacific isotope compositions should occur continuously along the length of the Pacific–Antarctic Ridge (PAR). Fresh basaltic glasses have now been recovered from the southernmost portion of this previously unsampled ridge axis, and we report their major element, trace element and isotopic compositions. The chemical systematics of these rocks suggest that the Pacific Ocean geochemical province includes the PAR, extends to the AAD south of Australia, and thus is one of the largest chemically coherent mantle domains on the Earth.

During a recent geophysical survey of a portion of the PAR, S. Cande, W. Haxby, C. Raymond and co-workers occupied two dredge sites on the ridge segments on either side of Fracture Zone XII at approximately 64° S, 171° W (Fig. 1 and Table 1), where the ridge is spreading at ~6 cm yr⁻¹ full-rate. Both dredges recovered basalts with abundant fresh glass. Before the collection of these samples, the only other basalt samples recovered from the southernmost Pacific were altered rocks dredged sig-

nificantly off-axis⁵ (see Fig. 1). Major element, trace element and isotope analyses of our PAR samples were performed on hand-picked basaltic glasses; the results and analytical techniques are presented in Table 1, with brief petrographic and site descriptions. All samples are of normal MORB (N-MORB) compositions and show limited intra-dredge variability⁶. Liquid line of descent calculations⁷ suggest that D1 is not related to the slightly more primitive D2 by simple crystallization from a common parental magma composition; notably, D1 is enriched by more than a factor of two in the most incompatible trace elements (for example, Ba and K) relative to D2. Samples from the PAR are remarkably similar in major and trace element composition to N-MORB samples from the northern East Pacific Rise (EPR) which have been analysed by the same technique^{8,9}. The most significant difference is that samples from the PAR appear to have higher SiO₂ than those from the northern EPR. In terms of global systematics, the PAR samples have Na_{8.0} and Fe_{8.0} values (Na and Fe normalized to 8 wt% MgO to minimize the effects of fractionation¹⁰), suggesting intermediate extents and pressures of melting, an observation consistent with their ridge depth of ~2,500 m (ref. 10).

Sr, Nd and Pb isotope analyses of these samples are shown in Fig. 2, and compared to fields for N-MORB from Atlantic, Pacific and Indian Ocean ridges. Samples from the two dredges are similar in isotopic composition (for example, ⁸⁷Sr/⁸⁶Sr = 0.70239–0.70242). Furthermore, all samples fall well within the fields for N-MORB from Pacific ridges, which include analyses from the northern and southern EPR as well as the Juan de Fuca and Gorda Ridges. It is important to note that although the fields for Pacific and Atlantic MORB overlap on plots of ²⁰⁶Pb/²⁰⁴Pb ratio against ²⁰⁸Pb/²⁰⁴Pb, ²⁰⁷Pb/²⁰⁴Pb and ⁸⁷Sr/⁸⁶Sr ratios, subtle differences exist between Atlantic and Pacific MORB as well^{11,12}. In particular, Pacific MORBs extend to lower ¹⁴³Nd/¹⁴⁴Nd ratios for the same ⁸⁷Sr/⁸⁶Sr ratio than Atlantic MORB (Fig. 2a). The PAR samples also show this subtle characteristic, and plot just outside the Atlantic MORB field and within the Pacific field. Thus, these results show that the PAR samples are isotopically distinct from both Indian and Atlantic Ocean MORB and similar to MORB recovered from eastern Pacific ridges thousands of kilometres from our study area.

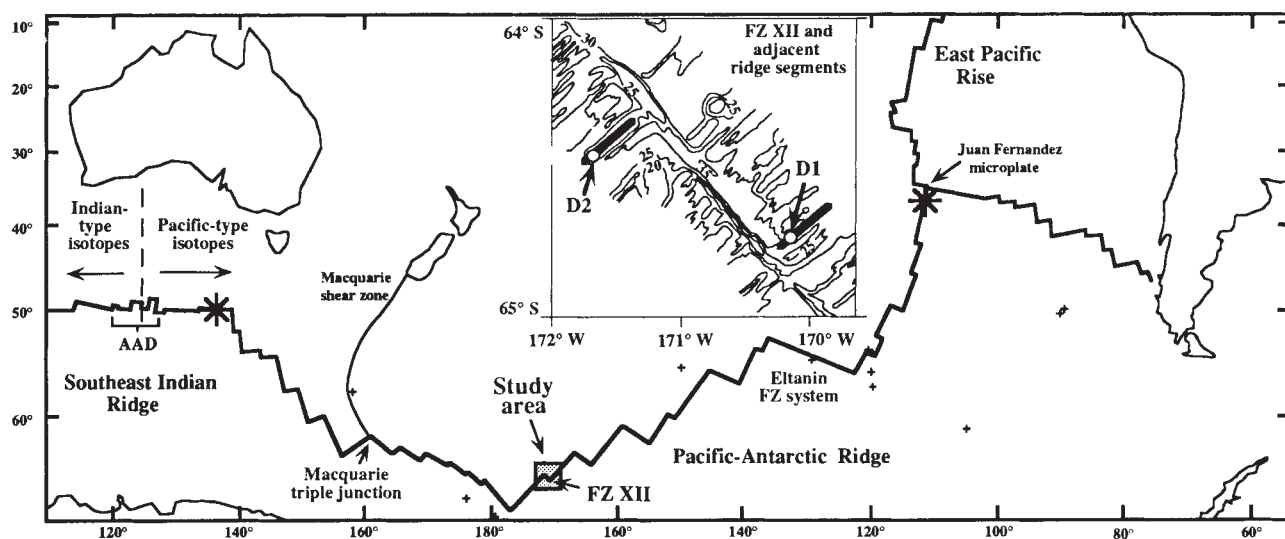


FIG. 1 Global system of oceanic spreading centres. The change in name from the Southeast Indian Ridge to the Pacific–Antarctic Ridge occurs at the Macquarie triple junction. Study area in the vicinity of Fracture Zone XII is shown by shaded box, and in detail in inset. Stars indicates locations of nearest other fresh samples recovered; ~4,000 km east of the study area near the Juan Fernandez microplate²⁵ and ~4,000 km

west of the study area near the Australian–Antarctic discordance along the Southeast Indian Ridge^{3,4,20,26}. Altered, off-axis samples analysed by Campsie *et al.*⁵ are shown as crosses. Inset: General bathymetry (S. Cande and W. Haxby, personal communication) of Fracture Zone XII; selected contours are labelled in hundreds of metres. Locations of ridge segments shown by solid lines and of two dredges by open circles.

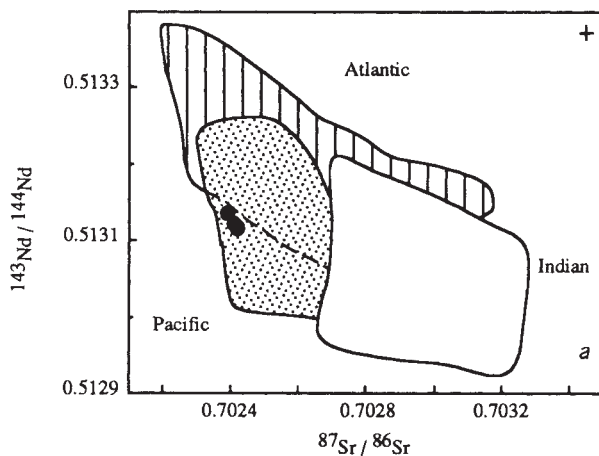
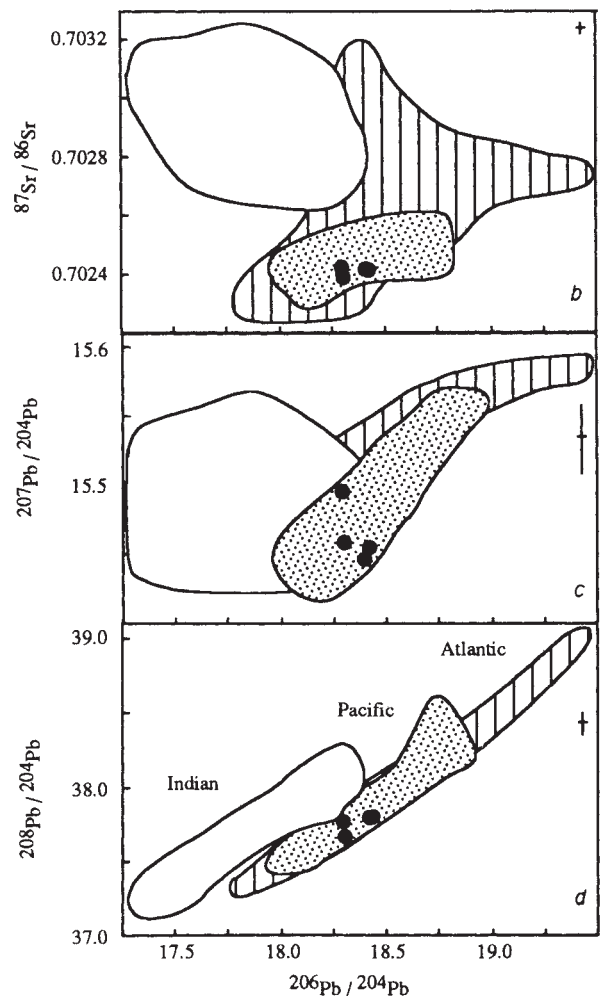


FIG. 2 Isotope ratios of basaltic glasses. Plot of $^{87}\text{Sr}/^{86}\text{Sr}$ against $^{143}\text{Nd}/^{144}\text{Nd}$ (a), and plots of $^{207}\text{Pb}/^{204}\text{Pb}$ against $^{87}\text{Sr}/^{86}\text{Sr}$ (b), $^{207}\text{Pb}/^{204}\text{Pb}$ (c), and $^{208}\text{Pb}/^{204}\text{Pb}$ (d). Data are shown by filled circles. Fields for N-MORB isotope data from North Atlantic (vertical ruled), Pacific (stippled) and Indian (open) Ocean ridges are shown for comparison. Analytical uncertainties are shown by crosses. References for MORB fields are as in Fig. 3 of Klein *et al.*³.



The growing data base on worldwide variations in MORB compositions has led to the identification of subtle differences in major and trace element systematics among the three ocean basin (for example, refs 10, 13–15). Such inter-ocean variations can be seen, for example, in $\text{Na}_{8.0}$ and $\text{Fe}_{8.0}$ systematics (Fig. 3); N-MORB from Pacific ridges tend to have lower $\text{Na}_{8.0}$ for the same $\text{Fe}_{8.0}$ (ref. 15). Our PAR samples share this general characteristic with other Pacific N-MORB. Thus, in addition to the isotope affinity to Pacific MORB noted above, these major element systematics suggest an affinity with N-MORB from eastern Pacific ridges, distinct from those from the Atlantic.

If we assume that the intervening unsampled portions of the EPR and PAR exhibit similar chemical characteristics, the length of ridge included in the 'Pacific MORB' province (including the Juan de Fuca Ridge) would extend southwest ~11,000 km to our study area, and on another 4,000 km to the Pacific-like samples east of the AAD (Fig. 1). If future sampling confirms our assumptions, this 15,000-km length of ridge would overlie one of the largest chemically coherent stretches of oceanic mantle on the Earth, second only in extent to the distinctive belt of ocean-island compositions centred on 30° S known as the 'Dupal anomaly'². The persistence of Pacific MORB-type mantle along

FIG. 3 Plot of $\text{Fe}_{8.0}$ against $\text{Na}_{8.0}$ (iron and sodium contents normalized to 8.0 wt% MgO to minimize fractionation effects)¹⁰ for basaltic glasses (filled circles). Fields for North Atlantic (vertical ruled) and Pacific (stippled) N-MORB, removed from the influence of hot spots, are shown for comparison. All data shown have $\text{K}_2\text{O}/\text{TiO}_2$ ratios of <0.15. The Pacific field consists of >200 analyses (fifteen outliers omitted) from diverse Pacific ridge segments; the North Atlantic field consists of >100 analyses (two outliers omitted) from 23°–30° N, the longest, well-sampled, portion of the northern Mid-Atlantic Ridge that is not influenced by a hot spot. Data from refs 17, 27, 28 and C. Langmuir (personal communication) and the Smithsonian Institution basalt glass data file (W. Melson, personal communication). Data analysed at the Smithsonian Institution had the following corrections applied to minimize inter-laboratory biases: $\text{Na}_2\text{O} \times 1.020$, $\text{MgO} \times 1.042$ and $\text{FeO} \times 0.976$. Also shown is the global trend of regional averages of $\text{Na}_{8.0}$ and $\text{Fe}_{8.0}$ (ref. 15), which in general represent data averaged over ~100 km of ridge length. $\text{Na}_{8.0} = \text{Na}_2\text{O} + 0.373(\text{MgO}) - 2.98$; $\text{Fe}_{8.0} = \text{FeO} + 1.664(\text{MgO}) - 13.313$; calculated for samples with 5.0–8.5 wt% MgO.

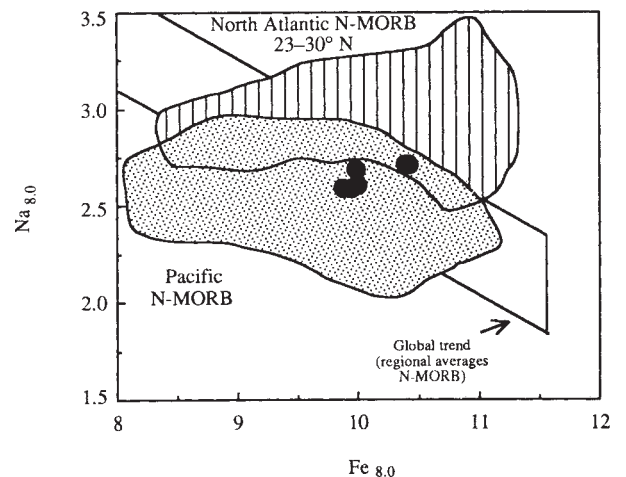


TABLE 1 Major and trace element and isotope analyses of basaltic glasses

	Dredge 1*			Dredge 2†		
	D1-S4	D1-S6	D1-S8-G1	D2-S1	D2-S4	D2-S7-G1
SiO ₂	50.47	50.00	50.14	50.63	50.66	50.75
Al ₂ O ₃	14.33	14.71	14.76	14.61	14.62	14.67
Fe ₂ O ₃	11.38	11.53	11.38	10.71	10.72	10.83
MgO	7.74	7.94	8.05	8.10	8.06	7.99
CaO	11.80	11.98	11.95	11.98	12.01	12.01
Na ₂ O	2.74	2.71	2.67	2.52	2.51	2.55
K ₂ O	0.09	0.09	0.09	0.05	0.05	0.05
TiO ₂	1.51	1.43	1.42	1.32	1.33	1.35
P ₂ O ₅	0.14	0.13	0.12	0.12	0.12	0.12
MnO	0.19	0.19	0.19	0.18	0.18	0.18
SUM	100.4	100.7	100.8	100.2	100.3	100.5
Ba	8.5	10.7	10.8	4.3	4.2	4.5
Sr	130	130	130	84	84	86
Zr	104	93	90	90	91	90
Y	33.8	32.0	32.1	31.8	32.1	31.7
V	307	278	283	288	283	305
Sc	40.9	40.2	41.0	40.7	40.5	41.5
Cu	68	75	73	71	88	73
Cr	200	285	294	268	255	265
Ni	66	79	86	84	78	79
²⁰⁶ Pb/ ²⁰⁴ Pb		18.296	18.289	18.418		18.395
²⁰⁷ Pb/ ²⁰⁴ Pb		15.457	15.494	15.454		15.446
²⁰⁸ Pb/ ²⁰⁴ Pb		37.670	37.775	37.809		37.782
⁸⁷ Sr/ ⁸⁶ Sr		0.70239 ± 2	0.70242 ± 2	0.70241 ± 2		0.70242 ± 2
¹⁴³ Nd/ ¹⁴⁴ Nd (oxide)		0.513135 ± 11	0.513118 ± 12	0.513122 ± 10		0.513116 ± 9
¹⁴³ Nd/ ¹⁴⁴ Nd (metal)			0.513115 ± 7			
Sr		110	110	94		94
Rb		0.86		0.33		
Nd (oxide)		9.44	8.97	9.19		10.55
Nd (metal)			8.99			
Sm		3.40	3.20	3.37		3.82

Analyses performed by d.c. plasma emission spectrometry at Lamont-Doherty Earth Observatory²⁰. Oxides reported in wt%; trace elements in p.p.m. Nd, Sr and Pb isotopic ratios and Nd, Sm, Rb and Sr abundances measured at the University of North Carolina-Duke mass spectrometry facility at Chapel Hill²¹⁻²⁴. Isotopic data are normalized to values of 0.11940 for ⁸⁶Sr/⁸⁸Sr, and 0.72190 for ¹⁴⁶Nd/¹⁴⁴Nd for both metal and oxide analyses. Measured replicate standard values are 0.710250 for ⁸⁷Sr/⁸⁶Sr for NBS-987, 0.512633 and 0.511851 for ¹⁴³Nd/¹⁴⁴Nd (as a metal) for BCR-1 and the LaJolla standard, respectively, and 0.512636 and 0.511852 for ¹⁴³Nd/¹⁴⁴Nd (as an oxide) for BCR-1 and the LaJolla standard, respectively. Pb data are referenced to the following values for NBS-981; 16.937, 15.491 and 36.721 for ²⁰⁶Pb/²⁰⁴Pb, ²⁰⁷Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb respectively. (Replicate standard analyses yielded a mean correction factor of 0.09% per amu.) Analytical uncertainties for Sr and Nd isotope ratios are $2\sigma/N^{1/2}$ (where σ is standard deviation and N is number of determinations), with values referring to the last decimal place(s) in the ratios. Maximum blank values for Sr, Nd and Pb are 0.3, 0.1 and 0.6 ng respectively.

* Dredge 1; 64° 44.4' S, 170° 8.6' W, Depth 2,429 m, Axial high, ~8 km northeast of Fracture Zone XII. Abundant olivine and plagioclase microphenocrysts.

† Dredge 2; 64° 26.4' S, 171° 36.9' W, Depth 2,518 m, Axial high, ~16 km southwest of Fracture Zone XII. ~5% plagioclase phenocrysts (2-3 mm), with plagioclase and olivine microphenocrysts.

the entire eastern boundary of the Pacific plate is consistent with the observation that the globe-encircling belt of the Dupal anomaly is, in fact, broken by non-Dupal ocean islands and non-Dupal-influenced ridges in the southeastern Pacific^{2,16}. Interestingly, one of the two Dupal maxima occurs in the central South Pacific Ocean basin^{2,16} so that the apparently continuous belt of Pacific MORB upper mantle, uncontaminated by these Dupal ocean island upwellings, skirts the perimeter of this anomaly.

In emphasizing the coherency of the Pacific MORB geochemical province, we do not mean to imply homogeneity of the mantle beneath circum-Pacific ridges. The sporadic occurrence of transitional MORB, enriched in both isotope ratios and incompatible trace elements abundances, is a fundamental characteristic of the EPR^{15,17}. Furthermore the extreme trace element variability observed along some segments of the Juan de Fuca Ridge¹⁸ attests unequivocally to the presence of intimate mixtures of depleted and more enriched components in the Pacific upper mantle. Nevertheless, among N-MORB from the world's ocean ridge system, there are assemblages of chemical characteristics that distinguish one region or ocean basin from another. These regional characteristics result from variations in the sizes and compositions of heterogeneous components in the

mantle, as well as chemical and physical factors that affect the style of melting, melt transport and crystallization processes. Our PAR data suggest simply that among the geochemical characteristics that can be explored with the limited number of sampling sites presently available, this central portion of the PAR shows chemical attributes with strong affinities to other Pacific ridges.

Clearly, with only two dredge sites over >8,000 km of ridge, a detailed comparison of the chemical systematics of the PAR with those observed along other ridge segments is not possible. There are numerous styles of geochemical variability that can be identified only with a high density of sampling sites over substantial lengths of ridge. Dense sampling of other mid-ocean-ridge segments has revealed, for example, that basalts from ridges with intermediate spreading rates, such as this portion of the PAR, can display either a positive correlation between Na_{8,0} and Fe_{8,0} (as found among samples from slow-spreading ridges) or an inverse correlation between these parameters (as observed on fast-spreading ridges^{13,15,19}). At present, only a small number of intermediate-spreading-rate ridges have been sampled in sufficient detail to identify these trends, and therefore the physical, chemical or geographic factors associated with each trend remain obscure. More extensive sampling is also necessary to

determine whether the higher SiO₂ contents of our PAR samples compared to EPR samples is simply part of the natural variability that exists in every region, or rather is a general and distinctive characteristic of the PAR^{10,13}. □

Received 29 July; accepted 28 September 1993.

- Dupré, B. & Allègre, C. J. *Nature* **303**, 142–146 (1983).
- Hart, S. R. *Nature* **309**, 753–757 (1984).
- Klein, E. M., Langmuir, C. H., Zindler, A., Staudigel, H. & Hamelin, B. *Nature* **333**, 623–629 (1988).
- Pyle, D. G., Christie, D. M. & Mahoney, J. J. *Earth planet. Sci. Lett.* **112**, 161–178 (1992).
- Campsie, J., Neumann, E. R. & Johnson, L. N. *J. Geol. Geophys. Res.* **26**, 31–45 (1983).
- Ferguson, E. M. et al. *Trans. Am. geophys. Un.* **73**, 605 (1992).
- Weaver, J. & Langmuir, C. *Computers Geosci.* **16**, 1–19 (1990).
- Langmuir, C. H., Bender, J. F. & Batiza, R. *Nature* **322**, 422–429 (1986).
- Bender, J. F., Langmuir, C. H. & Hanson, G. N. *J. Petrology* **25**, 213–254 (1984).
- Klein, E. M. & Langmuir, C. H. *J. geophys. Res.* **92**, 8089–8115 (1987).
- Ito, E., White, W. M. & Göpel, C. *Chem. Geol.* **62**, 157–176 (1987).
- White, W. M., Hofmann, A. W. & Puchelt, H. *J. geophys. Res.* **92**, 4881–4893 (1987).
- Klein, E. M. & Langmuir, C. H. *J. geophys. Res.* **94**, 4241–4252 (1989).
- Brodholt, J. P. & Batiza, R. *J. geophys. Res.* **94**, 4231–4239 (1989).

- Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges*, *Geophys. Monogr. No. 71* (eds Morgan, J. P., Blackman, D. K. & Sinton, J. M.) 183–280 (Am. Geophys. Un. Washington DC, 1992).
- Castillo, P. *Nature* **336**, 667–670 (1988).
- Sinton, J. M., Smaglik, S. M., Mahoney, J. J. & Macdonald, K. C. *J. geophys. Res.* **96**, 6133–6155 (1991).
- Karsten, J. L., Delaney, J. R., Rhodes, J. M. & Lias, R. A. *J. geophys. Res.* **95**, 19235–19256 (1990).
- Niu, Y. & Batiza, R. *J. geophys. Res.* **98**, 7887–7902 (1993).
- Klein, E. M., Langmuir, C. H. & Staudigel, H. *J. geophys. Res.* **96**, 2089–2107 (1991).
- Parrish, R. R., Roddick, J. C., Loveridge, W. D. & Sullivan, R. W. *Geol. Surv. Pap. Can.* **87**–(2), 3–7 (1987).
- Story, J. N. & Fritz, J. S. *Talanta* **21**, 892–894 (1974).
- Faris, J. P. & Warton, J. W. *Analyt. Chem.* **34**, 1077–1080 (1962).
- Strelow, F. W. E. *Analyt. Chem.* **32**, 1185–1188 (1960).
- Michael, P. J. & Chase, R. L. *Contr. Miner. Petrol.* **96**, 245–263 (1987).
- Anderson, R. N., Spariosu, D. J., Weissel, J. K. & Hayes, D. E. *J. geophys. Res.* **85**, 3883–3898 (1980).
- Rhodes, J. M., Morgan, C. & Lias, R. A. *J. geophys. Res.* **95**, 12713–12733 (1990).
- Davis, A. S. & Clague, D. A. *J. geophys. Res.* **92**, 10467–10483 (1987).

ACKNOWLEDGEMENTS. We thank S. Cande, W. Haxby, C. Raymond and the scientific party and crew of the RV Ewing (Leg EW9201) for the collection of these samples, and C. Langmuir, T. Plank, D. Desonie, J. Reynolds, D. Coleman, B. Beard and P. Curtis for their help with the chemical analyses and with the preparation of this manuscript. This work was supported by grants from the US National Science Foundation (E.K., S.C., W.H. and C.R.) and the Geological Society of America (E.F.).

Discrepancy between geological and geodetic deformation rates in the Ventura basin

Andrea Donnellan*, Bradford H. Hager† & Robert W. King†

* Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

† Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

SPACE geodesy provides a new tool for measuring neotectonic deformation which, when combined with geological and geophysical observations and models, can be used to infer rates and styles of deformation. In the Ventura basin region of southern California, the geodetically measured convergence rates of $\sim 7\text{--}10\text{ mm yr}^{-1}$ (ref. 1) are significantly lower than the $20\text{--}25\text{ mm yr}^{-1}$ estimated from geological data^{2,3}. Here we investigate the possible implications of this discrepancy. It might be, for example, that temporal variations cause the short- and long-term convergence rates to differ. But if this were the case we would expect to observe the long-term rate by extending the geodetic measurements over a broader region, yet only $11 \pm 3\text{ mm yr}^{-1}$ of regional convergence is observed⁴. Our preferred explanation of the discrepancy is that the geological models require revision, and to this end we present two deformation models that satisfy both the geological and geodetic constraints. Our modelling suggests that the faults bounding the basin are locked at the surface, but are slipping at depths below $\sim 2\text{--}5\text{ km}$.

To measure the continuing deformation in the Ventura basin region we collected four epochs of Global Positioning System (GPS) observations over a period of 4.6 yr. The relative velocities for most sites (with respect to site HOPP on the north side of the basin (Figs 1 and 2)), have been determined with an accuracy of better than 2 mm yr^{-1} at the 95% confidence level¹. The velocity field is dominated by convergence across the basin; the zone of shortening is $\sim 15\text{ km}$ wide and extends east–west for the 40 km extent of our network, with the convergence velocity increasing from east to west. The shear strain rate calculated in the region near Fillmore¹ is 60% greater than that measured across the San Andreas fault north of the basin⁵, and the dilata-

tional strain rate across the basin is unusually large. The angular strain rate in the region to the south of the basin is indistinguishable from zero at the 90% confidence level, whereas we detect a small amount of compression north of the basin and little strain in the Western Transverse Ranges northwest of the basin¹. The total contraction we measure across the Ranges south of the San Andreas fault ($11 \pm 2\text{ mm yr}^{-1}$) is lower than most geological estimates (that is, $17\text{--}26\text{ mm yr}^{-1}$ (ref. 6)). A similar rate is measured over a broader region⁴, implying that both the regional and the Ventura basin convergence rates estimated from the geology are too high.

There is evidence from both palaeomagnetic studies and geophysical studies that rotation, as well as strain, is an important component of deformation near the Ventura basin. Palaeomagnetic estimates of rotation rates near this region are uncertain, but range as high as 15° Myr^{-1} (ref. 7). The observation that the fault motion in major thrust earthquakes is nearly orthogonal to the direction of relative plate motion in the area is interpreted as evidence for the rotation of blocks of crust⁸. Our results indicate that these block rotations are continuing on timescales of a few years.

As a first approximation, we treat the region to the south of the basin, bounded by sites SCLA, HAPY, SAFE, CATO, and COTR, as a rigid block and calculate a weighted least-squares estimate of the motion of the block. The predicted velocities (Fig. 2) match the observed velocities to within their uncertainties. The rigid-body motion of the block can be expressed as a translation of its centre (point T in Fig. 2) at $4 \pm 1\text{ mm yr}^{-1}$ to the west and $6 \pm 1\text{ mm yr}^{-1}$ to the north, together with a clockwise rotation about its centre at $8 \pm 1^\circ\text{ Myr}^{-1}$. Equivalently, the motion relative to the north side of the basin can be described as a rotation about a pole located at point P near the eastern end of the basin. Viewed in this way, the region south of the basin appears as a nearly rigid block pivoting about an axis near its eastern end and squeezing the basin.

Our geodetic measurements of convergence rates suggest strongly that the geologically inferred rate of convergence across the Ventura basin is too high by a factor of about two. One possible explanation is that the assumed structures are correct, but that the geologic dates used to infer the rates are incorrect⁹. It is possible that the dates are a factor of two older than assumed (R. S. Yeats, personal communication). An alternative is that the inferred rates of slip on the faults bounding the basin are correct, but that the faults continue at the same dip angle into the lower crust, rather than flattening into décollements (Fig. 3) as is usually assumed¹⁰. We attempt to discriminate between these hypotheses by evaluating how well they fit the

observed pattern of deformation. Rather than using complicated three-dimensional fault models with laterally varying slip, we use simple two-dimensional elastic dislocation models driven by slip at depth on buried thrust faults (Figs 3 and 4), focusing on the east-central portion of the Ventura basin where the convergence rate is highest. The fault slip rates (assumed constant) are constrained so that the sum of their horizontal projections is

equal to the observed regional convergence rate of 11 mm yr^{-1} . Although there is no evidence of creep at the surface, the locking depths of these faults are unknown; we use models with the upper 2, 5 and 10 km locked.

The first model (Fig. 3) examines the surface deformation due to slip at depth on a proposed décollement on the south side of the basin¹¹, and on a similar structure extending beneath the

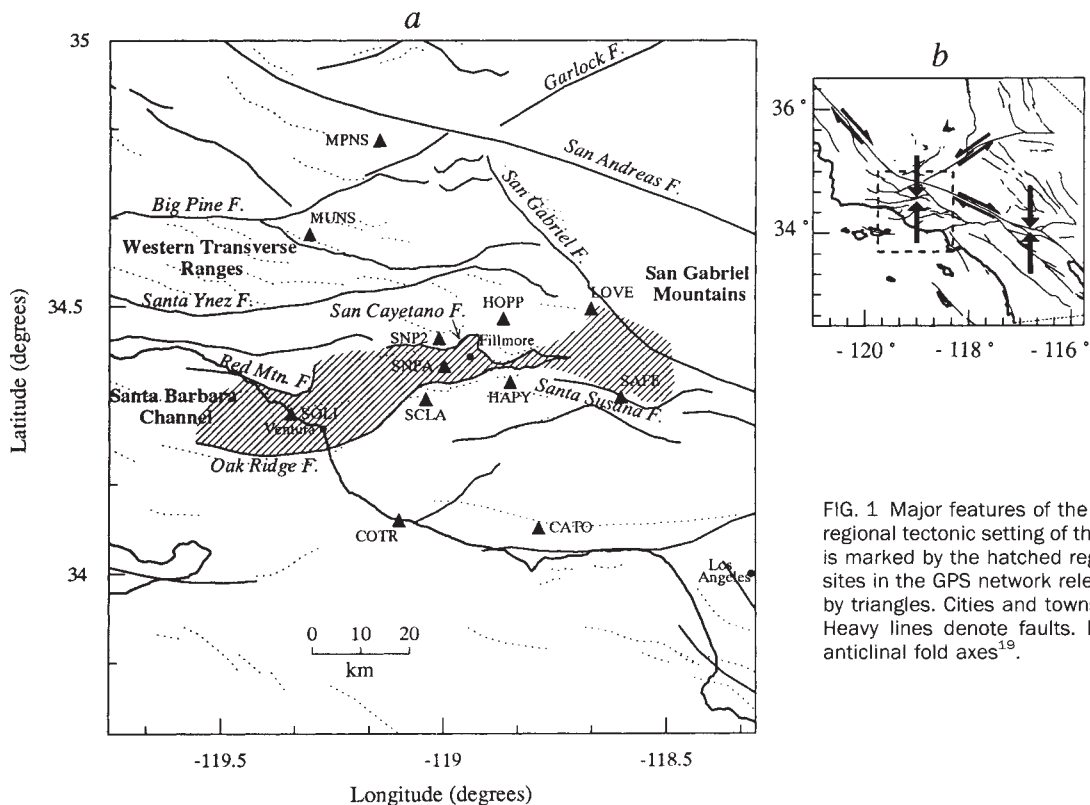
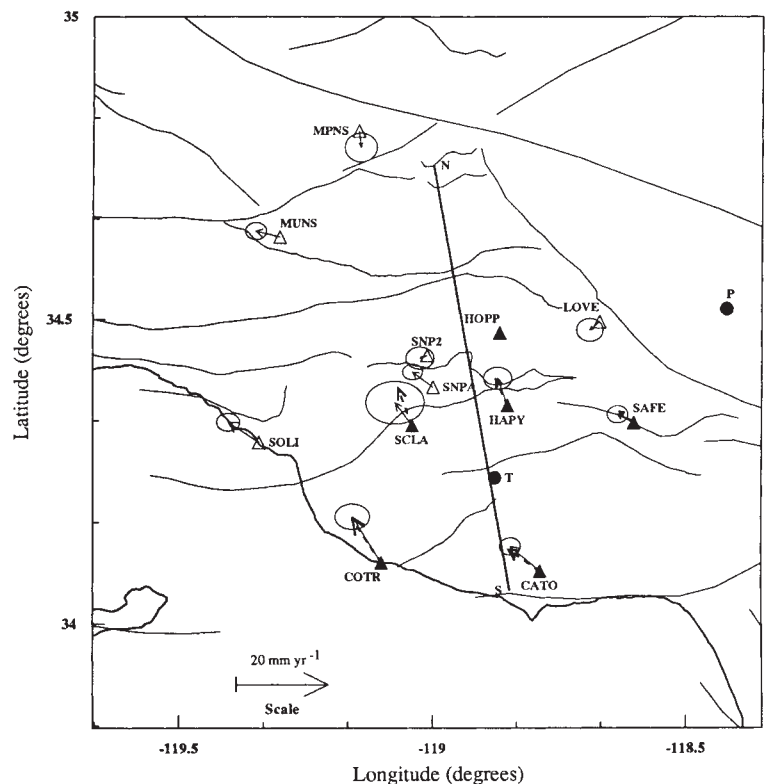


FIG. 1 Major features of the Ventura basin region (a), and regional tectonic setting of the basin (b). The Ventura basin is marked by the hatched region in a, and the locations of sites in the GPS network relevant to this study are marked by triangles. Cities and towns are marked by filled circles. Heavy lines denote faults. Dotted lines represent active anticlinal fold axes¹⁹.

FIG. 2 Velocities of sites relative to Hopper (HOPP) which is located slightly up and right from the centre of the Figure. The error ellipses represent 95% confidence. Sites on the southern block are marked by black triangles. T is the centroid for translation of the block. The dashed arrows are the calculated velocities for rotation of the block around point P.



San Cayetano fault on the north side¹². For locking depths of 2 and 5 km the model fits the data to within the 95% confidence error bars plotted. The depth of the possible décollement on the north side is not well constrained. A similar model, in which there is no northern décollement and the San Cayetano fault maintains a 45° dip at depth, also fits the data for slip deeper than 2–5 km.

The second model, which matches the inferred geologic rates on reverse faults that maintain their near-surface dips to great depth¹⁰, is also consistent with the geodetic data for locking depths of <10 km. Although our observations do not warrant fine-tuning of the model at this stage, relatively minor perturbations of the geometry of the structure or slip rates would allow us to fit the data even better. Numerical experiments involving variation of the depth and dip of slipping faults indicate that the geodetic data require thrust-faults that are slipping at shallow-to intermediate-depths both north and south of the basin. The

geology of the region suggests that the slipping sections terminate at locked thrust faults. The antisymmetry of the velocity profile across the basin also suggests that faults on both sides of the basin are responsible for the observed deformation rate.

In these models the faults extend infinitely along strike. Further west of the basin, in the Santa Barbara Channel, the geodetic rates of shortening^{13,14} are comparable to those observed in the Ventura basin, so a regional fault system that extends for at least 100 km is possible. If the buried slipping faults are in fact only 30 km long, they would have to be slipping at rates ~30% higher than our model suggests; in this case, however, we would have difficulty in fitting the regional convergence rate. We cannot yet distinguish the more likely fault structure with the current geodetic data, but to satisfy the observed velocity profile and be consistent with the geologic slip rates the faults must be slipping to depths shallower than 10 km. Oil well casings are not sheared at depths of 2.5–3 km so the locked

FIG. 3 Elastic dislocation models of the section N-S shown in Fig. 2. The relatively flat lying fault sections are décollements. *a*, Observed and modelled surface horizontal displacements relative to Hopper (HOPP). The solid line gives the displacement from the fault model shown, which has slip below 2 km depth. The dashed line is for a model with the same geometry, but with locked faults above 5 km depth. The dotted line is for locked faults above 10 km depth. The observed displacements and coordinates of the sites have been projected onto the section N-S. *b*, Cross-section along N-S showing the fault model that was used for the dislocation calculations. Slip rates are in mm yr^{-1} and are shown next to the fault segments.

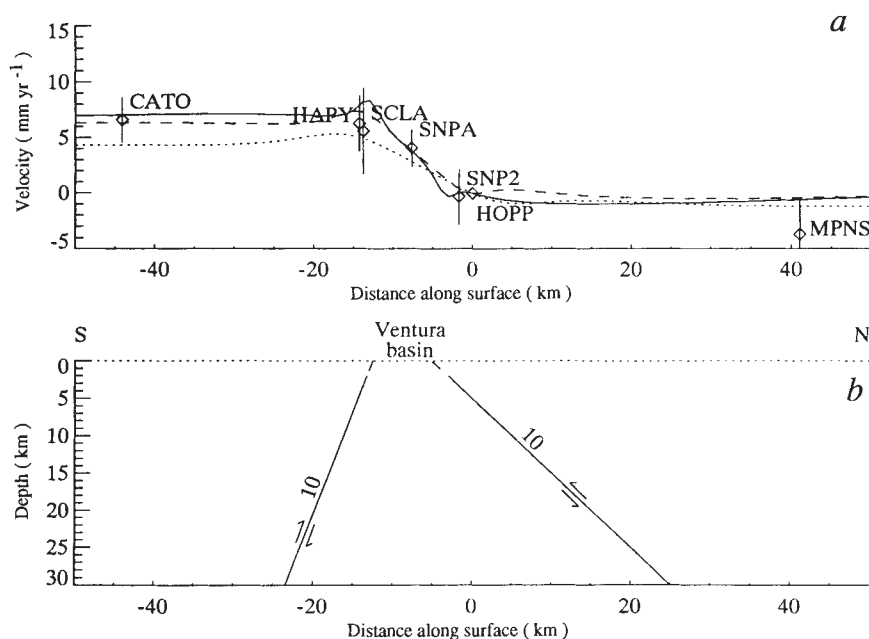
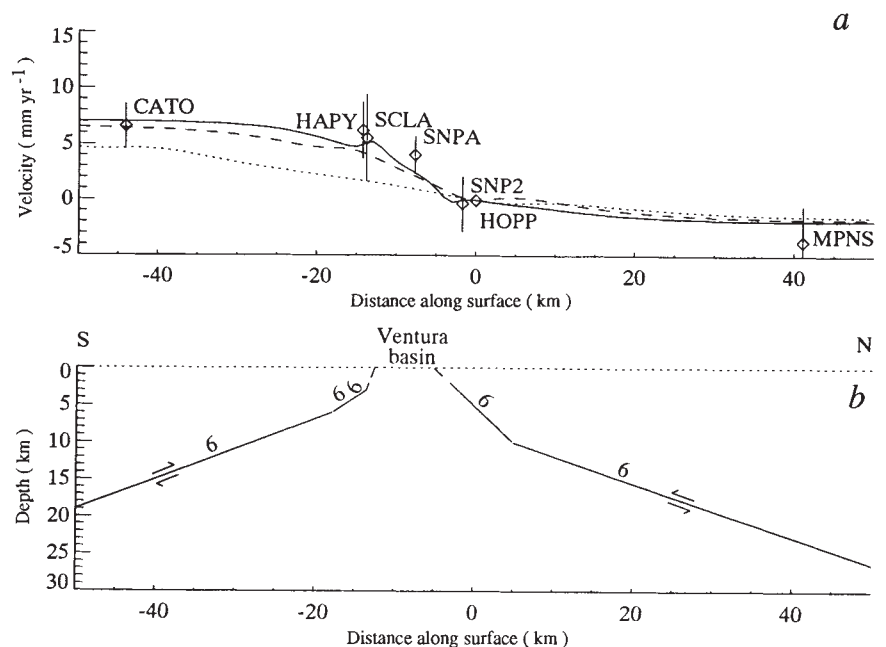


FIG. 4 Alternative models of the section N-S shown in Fig. 2 *a* and *b* are as described in Fig. 3 legend.

sections must extend deeper than this (R. S. Yeats, personal communication). Unusually low heat flow observed within the basin¹⁵ and anomalously deep earthquakes beneath the basin¹⁶ suggest substantial cooling due to downward advection of cool material, which is more compatible with the second model. We believe the deep earthquakes are due to loading of the basin and are unrelated to the faults slipping at the margins of the basin. More detailed future GPS measurements will place additional constraints on the extent and orientation of the sub-surface structures.

Slip on the modelled faults would cause strain to accumulate on the shallow locked sections—a process expected to result in earthquakes¹⁷. At present we do not know what fraction of the deformation associated with loading on these ramping structures is taken up seismically, as some of the deformation is clearly associated with active folding. A rupture dimension of 5×20 km, characteristic of the near-surface faults currently being loaded by the slipping sections, and an average stress drop of 60 bar (2 m of slip), would yield an earthquake of up to moment magnitude $M_w \approx 6.4$ (ref. 18). A moment magnitude $M_w \approx 7.5-8$ would result if the deep faults were locked rather than slipping. Using geology alone, it is difficult to determine which of these alternatives is more likely. Our space geodetic observations, combined with models of elastic strain accumulation, suggest that the deep faults are slipping, with only the near-surface accumulating strain, so a $M_w \approx 7.5-8$ event is unlikely. But the Santa Clara River Valley (within the Ventura basin) is densely popu-

lated, and a $M_w \approx 6$ earthquake is still large and potentially damaging. □

Received 22 April; accepted 4 October 1993.

- Donnellan, A., Hager, B. H., King, R. W. & Herring, T. A. *J. geophys. Res.* **98** (in the press).
- Yeats, R. S. *J. geophys. Res.* **88**, 569–583 (1983).
- Huflite, G. J. & Yeats, R. S. *J. geophys. Res.* (submitted).
- Feigl, K. et al. *J. geophys. Res.* **98** (in the press).
- Eberhart-Phillips, D., Lisowski, M. & Zoback, M. D. *J. geophys. Res.* **95**, 1139–1153 (1990).
- Namson, J. & Davis, T. *Geology* **16**, 675–679 (1988).
- Levi, S. & Yeats, R. S. *Tectonics* **12**, 688–702 (1993).
- Jackson, J. & Molnar, P. *J. geophys. Res.* **95**, 22073–22087 (1990).
- Molnar, P. Technical report (Southern California Earthquake Center, Univ. Southern California, Los Angeles, 1993).
- Yeats, R. S. & Huflite, G. J. *EOS* **73**, 590 (1992).
- Suppe, J. & Medwedeff, D. A. *Eclog. geol. Helv.* **83**, 409–454 (1990).
- Webb, T. H. & Kanamori, H. *Bull. seism. Soc. Am.* **75**, 737–757 (1985).
- Larsen, S. C., Agnew, D. C. & Hager, B. H. *J. geophys. Res.* **98**, 2119–2133 (1993).
- Kanamori, H. & Webb, F. H. *Geophys. Res. Lett.* **19**, 1491–1494 (1992).
- Larson, K. M. & Webb, F. H. *Geophys. Res. Lett.* **19**, 1491–1494 (1992).
- DeRito, R. F., Lachenbruch, A. H., Moses, T. H. & Munroe, R. J. *J. geophys. Res.* **94**, 681–699 (1989).
- Bryant, A. S. & Jones, L. M. *J. geophys. Res.* **97**, 437–447 (1992).
- Davis, T. L., Namson, J. & Yerkes, R. F. *J. geophys. Res.* **94**, 9644–9664 (1989).
- Kanamori, H. & Anderson, D. L. *Bull. seism. Soc. Am.* **65**, 1073–1096 (1975).
- Stein, R. S. & Yeats, R. S. *Sci. Am.* **260**(6), 48–57 (1989).

ACKNOWLEDGEMENTS. We thank everyone who assisted us with the GPS field observations; the Ohio State University, Scripps Institution of Oceanography and the University NAVSTAR Consortium (UNAVCO) who lent us receivers; the land owners, agencies and oil companies who allowed us access to private or restricted land; K. Feigl, D. Dong and T. Herring who contributed to the processing of the GPS data; and R. Yeats, J. Suppe and J. Shaw for comments on the manuscript. The research described in this paper was carried out under a National Research Council postdoctoral fellowship (to A.D.), at the NASA Goddard Space Flight Center, the Seismological Laboratory and Jet Propulsion Laboratory, California Institute of Technology, and the Massachusetts Institute of Technology under contracts with NASA, the US Geological Survey, and the US National Science Foundation.

Forest disequilibrium caused by rapid Little Ice Age cooling

Ian D. Campbell* & John H. McAndrews

Department of Botany, University of Toronto, and Department of Botany, Royal Ontario Museum, Toronto, Ontario, Canada M5S 2C6

GLOBAL climatic change may alter species' ranges as well as restructuring ecosystems^{1–3}. Models simulating forest growth predict that the area covered by different forest types may be affected², which may in turn further affect climate³. In the mixed forests of southern Ontario, pollen analyses have demonstrated that after AD 1400, *Fagus* (beech), the formerly dominant warmth-loving species, was replaced first by oak (*Quercus*) and subsequently by pine (*Pinus strobus*). Although these changes had been attributed to aboriginal forest clearance^{4,5}, they have also been seen in areas unaffected by aboriginal farming, and are now thought to reflect Little Ice Age cooling⁷. Although modelling suggests that some forests may take several centuries to reach equilibrium after a climatic change^{8,9}, a real forest showing this behaviour has not previously been identified. Here we model the Little Ice Age by a 2 °C decrease in mean annual temperature from AD 1200 to 1850, and show that the changes predicted by a forest simulator derived from FORET¹⁰ match those seen in southern Ontario. These forests thus appear to have remained in disequilibrium with the prevailing climate for more than 650 years.

The FORET-derived model uses an annual time step in which individuals (of 52 species in this case) are allowed to grow, compete and die. Growth is constrained by species-specific responses to aspects of the environment (such as growing degree-days or the absence of exposed mineral soil for seedbed), limits of the species itself (such as maximum growth rate or maximum height), and by competition with each other (modelled primarily as mutual shading). FORET and similar models have been used with great success in a wide variety of applications, including

simulating the effects of a pathogen, post-glacial forest development and the dynamics of several forest types around the world¹⁰. This simulated succession is converted to a pollen diagram through regression equations relating the relative abundance of each tree species to the relative abundance of its pollen^{11–13} (Fig. 3). The converted simulation results mimic actual pollen records from the region¹⁴. Hence we can use this

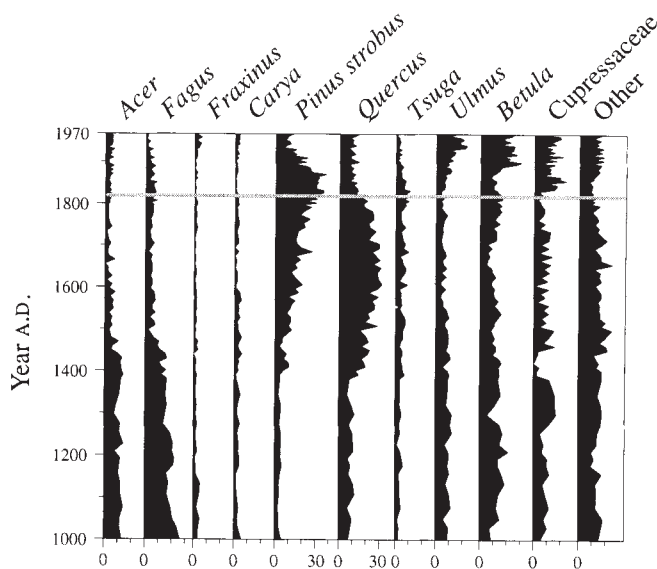
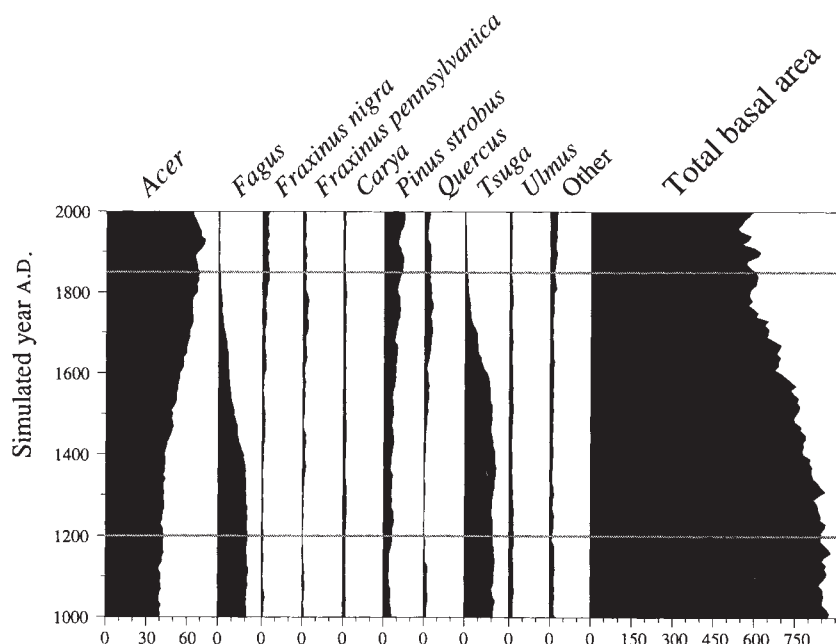


FIG. 1 Relative abundance pollen diagram of selected taxa from Crawford Lake, the most detailed pollen diagram of the last millennium available for southern Ontario. A rapid decline in *Fagus* begins ~AD 1450, leading to a succession of first *Quercus* then *Pinus strobus*. The horizontal line after AD 1800 indicates the first signs of European disturbance. The horizontal axis is per cent of total pollen (10 per cent per unit).

* Present address: Forestry Canada, 5320–122 Street, Edmonton, Alberta, Canada T6H 3S5.

FIG. 2 Simulated forest succession shows a decrease in *Fagus grandifolia* starting immediately with the onset of cooling; it is followed by increases in *Quercus rubra*, *Pinus strobus*, *Fraxinus nigra* and *Acer saccharum*; *Tsuga canadensis* declines later than *Fagus*. The total basal area also declines, indicating a more open forest containing less biomass. The simulation climate was monotonically cooled by 2 °C from AD 1200 to 1850, as suggested by proxy data for the Little Ice Age from nearby Michigan²³. The horizontal axis is per cent basal area (10 per cent per unit) except for the total basal area, which is m^2ha^{-1} .



simulation to examine the dynamics of this period of southern Ontario's forest history.

The most important feature of the simulation results is the decline in total basal area (Fig. 2), initiated immediately after the start of cooling (~AD 1200). By AD 1850, basal area reaches a value of less than 70% of the pre-cooling mean, implying at least a one-third reduction in total biomass per hectare. It is significant that this decrease starts at the same time as the first decline in *Fagus* abundance, suggesting that the premature death of dominant *Fagus* trees, rather than a failure of *Fagus* regeneration (as saplings of other species would take the place of the *Fagus* saplings), is initially implicated in this basal area decrease.

Furthermore, although *Quercus* and *Pinus strobus* do increase dramatically in both the real and simulated pollen records, this is due largely to their strong pollen production and dispersal; their simulated basal areas increase only slightly. Rather, the co-dominant *Acer saccharum*, which produces only a moderate quantity of pollen, is the species which eventually most benefits from the decline in *Fagus* in the simulation.

During the Little Ice Age, increased canopy opening after the death of *Fagus* would have allowed *Quercus*, *Fraxinus* and *Pinus strobus* to expand, and eventually reduced competition for *Acer saccharum*. *Tsuga canadensis*, however, reproduces poorly under an open canopy¹⁵, and its eventual decline may result from canopy opening as much as from climate cooling. Thus the sequence of events observed in the pollen records of this region was essentially a combination of seral and climatic successions initiated by a rapid climate cooling.

In this simulation, climate cooling was stopped at AD 1850. The forest succession continues, however, to the end of the simulation run at AD 2000, and presumably would continue beyond, so that even 150 years after the end of the climate cooling the forest would not yet have reached equilibrium with the prevailing climate.

In the actual pollen record the succession is truncated by logging and land clearance in the mid-1800s. Although the Little Ice Age was coming to an end at this time, it seems likely that the forest had not yet reached equilibrium with the climate, as the model suggests seral succession would have continued in the absence of further cooling. Although many ecologists, palaeo-ecologists, and park managers use the pre-clearance vegetation as their baseline 'equilibrium' forest, our results show that as early as AD 1450, the forest composition and structure were in disequilibrium with the prevailing climate. Although many

authors have considered the possibility of a lag between change and forest response, due to the time needed for soil development¹⁶, species migration¹⁷, or simply for mature stems to die and be replaced by different species¹⁸, few have considered the possibility of a rapid climate change triggering a seral type of succession.

Although a similar forest simulator was used to demonstrate the possibility of a disequilibrium forest resulting from Little Ice Age cooling⁸, such a forest has not previously been identified, and remained until now a theoretical construct.

Although the wild forest of southern Ontario, now largely cleared, is no longer a major carbon reserve, disequilibrium dynamics have been simulated for global warming scenarios in other forest types¹⁹. The implications for future climate change are significant. First, a transient decrease in forest biomass may result from rapid climate change. Tree death on a massive scale

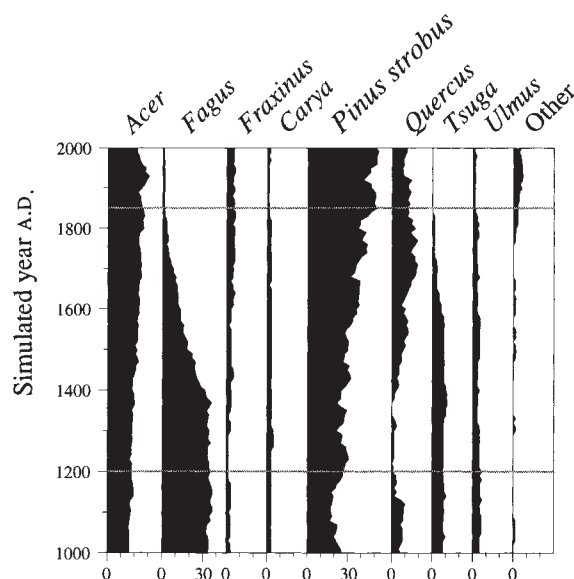


FIG. 3 Simulated pollen record created by applying pollen/biomass regression equations to the simulated forest succession. The simulated pollen record strongly resembles actual pollen records from the region¹⁴. The horizontal axis is per cent pollen (10 per cent per unit).

would produce a large fuel supply for wildfire, and would in any event result in a flux of stored carbon into the atmosphere. The opening of the forest canopy would also increase the albedo of the landscape, leading to cooling of the land surface³. Either effect has serious implications for future climate change. Furthermore, although there is still some dispute over the potential migration rates of tree species in response to climate change, forest equilibrium is not only a matter of species presence, but also of forest structure^{20,21}. Even with all relevant species and suitable soils present, it may nevertheless take several centuries for a forest to equilibrate with climate because of the lags inherent in seral succession; this has been simulated for northern Michigan mixed forest under future climate change scenarios⁹.

In other regions, a succession such as this might not be observed in the pollen record. In southern Ontario, its prominence is due to several factors, including the high pollen representation of some of the species involved in the succession, the ecotonal nature of the region, the long lifespans of many of the tree species, and the infrequent prehistoric fire regime²² which allowed a long seral succession to develop. Although a similar disequilibrium probably occurred in other regions, the absence of any one of these factors would render it nearly unobservable in the pollen record. □

- The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 283–310 (Cambridge University Press, New York, 1990).
2. Pastor, J. & Post, W. M. *Nature* **334**, 55–58 (1988).
 3. Bonan, G. B., Pollard, D. & Thompson, S. L. *Nature* **359**, 716–718 (1992).
 4. McAndrews, J. H. in *Vegetation History* (eds Huntley, B. & Webb, T. III) 673–697 (Kluwer, New York, 1988).
 5. McAndrews, J. H. & Boyko-Diakonow, M. in *Quaternary Geology of Canada and Greenland* (ed. Fulton, R. J.) 528–530 (Geological Survey of Canada, Ottawa, 1989).
 6. Burden, E. T., McAndrews, J. H. & Norris, G. *Can. J. Earth Sci.* **23**, 43–54 (1986).
 7. Campbell, I. D. & McAndrews, J. H. *Can. J. Bot.* **69**, 1719–1730 (1991).
 8. Davis, M. B. & Botkin, D. B. *Quat. Res.* **23**, 327–340 (1985).
 9. Solomon, A. M. & Bartlein, P. J. *Can. J. For. Res.* **22**, 1727–1738 (1992).
 10. Shugart, H. H. *A Theory of Forest Dynamics* (Springer, New York, 1984).
 11. Davis, M. B. *Am. J. Sci.* **261**, 897–912 (1963).
 12. Delcourt, P. A., Delcourt, H. R. & Webb, T. III *Atlas of Mapped Distributions of Dominance and Modern Pollen Percentages for Important Tree Taxa of Eastern North America* (Am. Ass. Stratigraphic Palynologists, Dallas, 1984).
 13. Bradshaw, R. H. W. & Webb, T. III *Ecology* **66**, 721–737 (1985).
 14. Campbell, I. D. thesis, Univ. Toronto, 1992.
 15. Godman, R. M. & Lancaster, K. in *Silvics of North America vol. 1: Conifers* (eds Burns, R. M. & Honkala, B. H.) 604–612 (United States Department of Agriculture Forest Service Handbook 654, Washington DC, 1990).
 16. Pennington, W. *Vegetatio* **67**, 105–118 (1986).
 17. Davis, M. B. *Ecology* **50**, 312–332 (1969).
 18. Bryson, R. A. *Quat. Res.* **15**, 143–159 (1981).
 19. Solomon, A. M. *Oecologia* **68**, 567–579 (1986).
 20. Davis, M. B. in *Community Ecology* (eds Diamond, J. & Case, T. J.) 269–284 (Harper and Row, New York, 1986).
 21. Delcourt, H. R., Delcourt, P. A. & Webb, T. III *Quat. Sci. Rev.* **1**, 153–175 (1983).
 22. Johnson, E. A. *Fire and Vegetation Dynamics: Studies from the North American Boreal Forest* (Cambridge Univ. Press, New York, 1992).
 23. Bernabo, J. C. *Quat. Res.* **15**, 143–159 (1981).

Received 5 April; accepted 8 September 1993.

1. Melillo, J. M., Callaghan, T. V., Woodward, F. I., Salati, E. & Sinha, S. K. in *Climate Change:*

ACKNOWLEDGEMENTS. We thank M. Apps, E. Hogg and S. Zoltai for comments on the manuscript.

Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*

U. Hentschel & H. Felbeck

Marine Biology Research Division 0202, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0202, USA

THE vestimentiferan tubeworm *Riftia pachyptila* is found around hydrothermal vent areas in the deep sea. Intracellular bacterial chemoautotrophic symbionts use the oxidation of sulphide from the effluent of the vents as an energy source for CO₂ fixation. They apparently provide most or all of the nutritional requirements for their gutless hosts^{1–5}. This kind of symbiosis has since been found in many other species from various other phyla from other habitats^{6–9}. Here we present results that the bacteria of *R. pachyptila* may cover a significant fraction of their respiratory needs by the use of nitrate in addition to oxygen. Nitrate is reduced to nitrite, which may be the end product (nitrate respiration)¹⁰ or it may be further reduced to nitrogen gas (denitrification)¹¹. This metabolic trait may have an important role in the colonization of hypoxic habitats in general by animals with this kind of symbiosis.

Because the symbionts cannot be cultured¹² we used physically purified bacteria¹³ for our experiments which were incubated under anaerobic conditions in the presence of nitrate. Nitrite appeared in the medium at a linear rate (0.91 ± 0.07 nmol per mg protein per min, $n=3$) for at least 90 min (Fig. 1) and was independent of the nitrate concentration between 50 μ M and 1 mM. Nitrite was not formed without nitrate ($n=3$), with plume tissue containing no symbionts ($n=1$), when the bacteria were boiled ($n=3$), or when cyanide was added ($n=2$). When the concentration of nitrate dropped below 20 μ M (Fig. 1), nitrite was reduced as well, indicating that nitrite can be respired further. There was no indication that the symbionts can respire nitrate to ammonium. Nitrate respiration was stimulated by sulphide with maximal rates at 500 μ M sulphide (Fig. 2), but not by thiosulphate. This supports previous results^{14,15} that the sym-

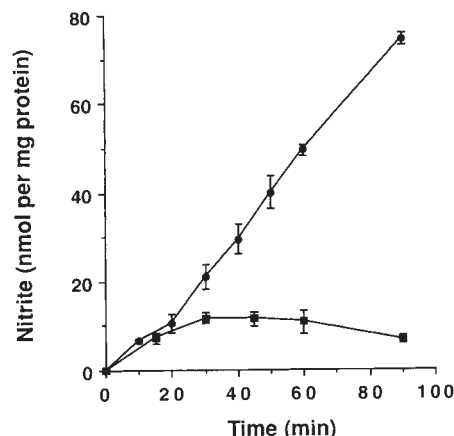


FIG. 1 Production of nitrite by purified symbionts from *R. pachyptila*. Live tubeworms were collected in temperature-insulated containers from the Genesis vent site at 2,638 m depth at 12° 48.675' N, 103° 56.386' W on the East Pacific Rise. Symbionts were purified under anaerobic conditions¹³ from crude homogenates of trophosome¹⁷, the tissue containing the bacteria. Isolated symbionts were incubated anaerobically in 10 ml *Riftia* saline²⁴ containing 500 μ M sulphide and 1 mM (●) or 50 μ M nitrate (■). The concentration was adjusted to about 2 mg symbiont protein ml⁻¹. The incubations were done at 15 °C and stirred continuously. To maintain anaerobic conditions, each incubation chamber was closed with a rubber septum and a nitrogen gas stream was introduced through a needle 15 min before and during the experiment. At 10 min intervals, 1.5 ml aliquots were drawn through a syringe and centrifuged for 5 min in a Fisher table-top centrifuge. The concentration of nitrite in the supernatant of the samples (nmol per mg symbiont protein) was determined after removal of sulphide²⁵. Data points represent means $\pm$ s.e., $n=3$.

bionts are sulphide specialists.

Because denitrification is usually a trait found in bacteria from anaerobic environments, the influence of oxygen on the reduction of nitrate by the symbionts had to be assessed. The only measurements of oxygen consumption by purified *R. pachyptila*

symbionts have been made at oxygen saturation ($260 \mu\text{M}$)¹⁵. But these measurements appear to be unrealistic, because the levels of free oxygen in the bacteriocytes, the cells containing the symbionts, will be much lower. They can be roughly estimated from the physiology of the haemoglobin dissolved in the blood. Oxygen is delivered from the plume to the trophosome through the vascular blood in a closed circulatory system. The trophosome itself is completely submerged in the coelomic fluid filling the coelomic cavity. Both, the vascular and coelomic haemoglobin molecules have an extremely high affinity for oxygen ($P_{50}=0.1-0.3$ Torr at pH 7 (ref. 16) equivalent to about $2 \mu\text{M}$). Because the concentration of free oxygen inside the tissue is determined by the PO_2 of the blood, its concentration in the bacteriocytes should be maximally around $2 \mu\text{M}$ (ref. 17). This is a conservative estimate, because factors such as oxygen consumption in the cytosol or diffusion barriers could lower the free oxygen levels even more. The large pool of oxygen bound to haemoglobin¹⁸ should not increase the concentration of free oxygen in the bacteriocyte cell but only maintain it at a low level. It has been proposed^{6,17} that one of the major roles of the two haemoglobins is to decrease the concentration of free oxygen in the immediate vicinity of the symbionts possibly resulting in a reduction of the oxygenase reaction of the symbionts' ribulose 1,5-bisphosphate carboxylase, the key enzyme of the Calvin-Benson cycle.

Although oxygen respiration decreases with decreasing oxygen concentrations, nitrate respiration increases (Fig. 3). At 100% saturation, oxygen respiration is 20 times higher than nitrate respiration. At the lowest oxygen concentration (below 10% saturation equal to $26 \mu\text{M}$ or less) the rates are statistically equivalent. We did not determine oxygen consumption at lower levels because they could not be measured reliably with the techniques available to us. We can calculate that under estimated *in vivo* conditions (at a free oxygen concentration of $2 \mu\text{M}$ or less and availability of nitrate in a concentration range similar to that of sea water) at least half of the oxidative capacity of the symbionts can be provided by nitrate. This is assuming no enhancement in respiration or increase in oxygen concentration in the bacteriocytes by the blood binding capacity for oxygen¹⁸.

FIG. 3 Oxygen and nitrate respiration rates at different oxygen concentrations. Black bars represent oxygen respiration rates, grey bars are nitrate respiration rates (mean $\pm$ s.e., n given above each bar).

METHODS. Symbiont oxygen respiration was measured with a modified Clarke-type oxygen electrode²⁶. Before every incubation, nitrogen gas was bubbled in the incubation mixture through small-diameter tubing inserted into the side groove of the electrode setup until the desired oxygen concentration was established. Symbiont suspension was added to a final concentration of $1-2 \text{ mg ml}^{-1}$ in 1.9 ml *Riftia* saline²⁴ and allowed to equilibrate for several min. The oxygen consumption rate under these conditions was defined as the baseline. Then sulphide was added to a final concentration of $200 \mu\text{M}$ and the sulphide-stimulated oxygen consumption rate measured. After a few minutes oxygen was reintroduced to reestablish the concentration when sulphide had been added. Then azide was added (final concentration $50 \mu\text{M}$) as a respiratory inhibitor. The azide-insensitive oxygen consumption rate was measured and found to be largely independent of the sulphide or oxygen concentrations in the range used. To avoid the introduction of artefacts due to the order in which the rates were measured we changed the oxygen concentrations randomly before we measured nitrate and oxygen consumption, the results were identical. The oxygen respiration rate at 90–70% saturation was measured at the beginning and at the end of each preparation to ensure the viability of the symbiont suspension. They remained constant for several hours, when the symbiont suspension was kept on ice. Nitrate respiration rates

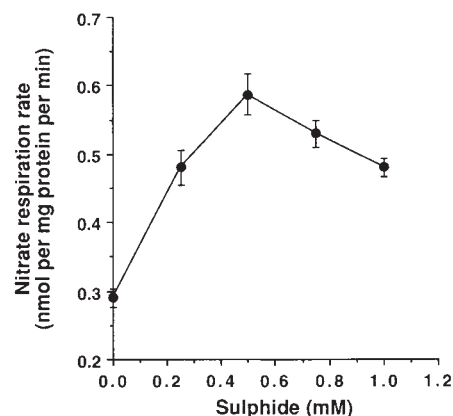
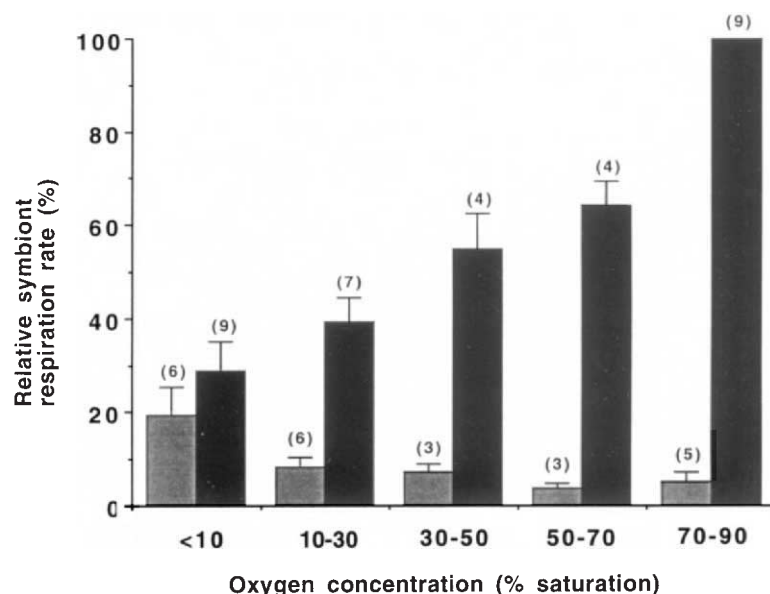


FIG. 2 Influence of sulphide on nitrate respiration in purified symbionts. Incubations were done in *Riftia* saline²⁴ with sulphide concentrations between 0 and 1 mM nitrate. At timepoints, the concentrations of nitrite were determined and the respiration rate ($\text{nmol nitrite mg}^{-1} \text{ min}^{-1}$) calculated. Data points represent means $\pm$ s.e., $n=4$.

This use of nitrate for respiration allows the oxygen around the symbionts to be maintained at extremely low concentrations yet energy can still be gained through respiratory pathways.

Nitrate respiration can also be measured in the intact symbiosis. In preliminary experiments small animals (1–4 cm long) were incubated under pressure in closed vials in sea water containing 1 mM nitrate. They excreted nitrite at rates of up to 250 nmol per g worm fresh weight per h. These rates are corrected for activities of epibiotic bacteria on the tubes by removing the worm from its tube at the end of the incubation and continuing to measure nitrite production by the empty tube. But during these experiments the animals were stressed. The result is, therefore, only indicative of their ability to reduce nitrate, not of their *in situ* metabolic rate.

In the hydrothermal vent environment, both the presence of



were measured after addition of 1 mM nitrate. After 15 min, an aliquot was taken, centrifuged for 5 min and the concentration of nitrite determined in the supernatant. The rates are expressed relative to the sulphide-stimulated oxygen respiration rate at oxygen saturation.

nitrate and its consumption near tubeworm clusters have been demonstrated¹⁹. The rate of nitrate removal was five times higher than necessary if it were due to nitrate assimilation. In this area, characterized by silicate concentrations of 500–600 μM , the animals can be exposed to conditions where the concentration of oxygen is only double that of nitrate (around 30 μM). The sulphide consumption rate of tubeworms *in situ* was highest under these conditions¹⁹.

In addition, initial data indicate that the symbionts of the clam *Lucinoma aequizonata*, collected in a hypoxic basin from the coast of Santa Barbara (CA), may also have the ability to use nitrate respiration to supplement the lack of oxygen²⁰.

If nitrate respiration were to be of major significance for the tubeworms then a means to transport nitrate to the symbionts inside the trophosome has to be provided. The concentration of nitrate in the blood could not be determined accurately, because it coprecipitated on denaturation of the blood proteins. Several methods used for the determination of nitrate in mammalian blood gave erroneous results in *R. pachyptila* blood^{21,22}. Nitrite, by contrast, was measured both in the vascular ($31.6 \pm 6.5 \mu\text{M}$, $n=11$) and the coelomic blood ($3.4 \pm 1 \mu\text{M}$, $n=12$) of freshly collected animals. Because the concentration of nitrite in the ambient water around the vent sites is, as commonly found in sea water, below 1 μM (K. Johnson, personal communication), it must originate from the symbionts' respiration. This result appears to be in contrast with others showing an even distribution of small molecules between vascular and coelomic fluid²³. In our experiments, the vascular blood samples used for nitrite determination were taken from the dorsal trunk vessel, the vessel carrying blood from the trophosome to the plume. It is, therefore, feasible that nitrite is released into the vascular blood stream in the trophosome and efficiently excreted in the plume.

Although the average concentration of nitrite in the vascular blood was 31.6 μM , three samples had much higher nitrite concentrations of 89 μM , 138 μM and 218 μM (values not included

in the average). Apparently, certain conditions promote higher concentrations of nitrite and presumably nitrate in the blood. Further studies are necessary to determine how nitrate is transported, whether specific uptake systems exist, and whether nitrate can be concentrated and bound in the blood. □

Received 24 May; accepted 17 September 1993.

1. Felbeck, H. *Science* **213**, 336–338 (1981).
2. Felbeck, H., Childress, J. J. & Somero, G. N. *Nature* **293**, 291–293 (1981).
3. Corliss, J. B. et al. *Science* **203**, 1073–1083 (1979).
4. Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. & Waterbury, J. B. *Science* **213**, 340–342 (1981).
5. Bosch, C. & Grasse, P. P. C. R. Acad. Sci. Paris (ser III) **299**, 413–419 (1984).
6. Childress, J. J. & Fisher, C. R. *Oceanogr. Mar. Biol. A. Rev.* **30**, 337–441 (1992).
7. Felbeck, H. in *Endocytobiology IV* (eds Nardon, P., Gianinazzi-Pearson, V., Grenier, A. M., Margulis, L. & Smith, D. C.) 327–334 (Institut National de la Recherche-Agronomique, Paris, 1990).
8. Fisher, C. R. *Rev. Aquat. Sci.* **2**, 399–436 (1990).
9. Tunnicliffe, V. *Palaeos* **7**, 338–350 (1992).
10. Stewart, V. *Microbiol. Rev.* **52**, 190–232 (1988).
11. Tiedje, J. M. in *Biology of Anaerobic Microorganisms* (ed. Zehnder, A. J. B.) 179–244 (Wiley, New York, 1988).
12. Distel, D. L. & Wood, A. P. *J. Bact.* **174**, 6317–6320 (1992).
13. Distel, D. L. & Felbeck, H. *J. exp. Zool.* **247**, 11–22 (1988).
14. Belkin, S., Nelson, D. C. & Jannasch, H. W. *Biol. Bull.* **170**, 110–121 (1986).
15. Wilmet, D. B. J. & Vetter, R. D. *Mar. Biol.* **106**, 273–284 (1990).
16. Arp, A. J. & Childress, J. J. *Science* **213**, 342–344 (1981).
17. Fisher, C. R., Childress, J. J. & Minnich, E. *Biol. Bull.* **177**, 372–385 (1989).
18. Childress, J. J., Arp, A. J. & Fisher, C. R. *Mar. Biol.* **83**, 109–124 (1984).
19. Johnson, K. S., Childress, J. J., Hessler, R. R., Sakamoto, A. C. M. & Beehler, C. L. *Deep Sea Res. Part A Oceanogr. Res. Pap.* **35**, 1723–1744 (1988).
20. Hentschel, U., Cary, S. C. & Felbeck, H. *Mar. Ecol. Prog. Ser.* **94**, 35–41 (1993).
21. Green, L. C. et al. *Analyt. Biochem.* **126**, 131–138 (1982).
22. Cortas, N. K. & Wakid, N. K. *Clin. Chem.* **36**, 1440–1443 (1990).
23. Childress, J. J. et al. *Biol. Bull.* **180**, 135–153 (1991).
24. Fisher, C. R., Childress, J. J. & Sanders, N. K. *Symbiosis* **5**, 229–246 (1988).
25. Gieskes, J. & Peretsman, *Ocean Drilling Program Tech. Note* 5, 30–36 (1986).
26. O'Brien, J. & Vetter, R. D. *J. exp. Biol.* **149**, 133–148 (1990).

ACKNOWLEDGEMENTS. We thank the captains and crew members of the RV *John Vickers*, RV *Nadir*, RV *Atlantis II* and the pilots and support crews of the submarines *Nautille* and *Alvin* for their assistance, J. J. Childress, C. R. Fisher and R. Lee for helpful discussions, D. Oros for assistance with the whole-animal experiments and P. Turner for cruise organization and assistance. This research was supported by NSF and a travel grant from NATO.

The retinoid ligand 4-oxo-retinoic acid is a highly active modulator of positional specification

W. W. M. Pijnappel^{*†}, H. F. J. Hendriks[†],
G. E. Folkers^{*}, C. E. van den Brink^{*}, E. J. Dekker^{*},
C. Edelenbosch^{*}, P. T. van der Saag^{*}
& A. J. Durston^{*}

^{*}Hubrecht Laboratory, Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
[†]TNO Institute of Ageing and Vascular Research (IVVO), PO Box 430, 2300 AK, Leiden, The Netherlands

RETINOIDS (vitamin A and its metabolites) are suspected of regulating diverse aspects of growth, differentiation, and patterning during embryogenesis¹, but many questions remain about the identities and functions of the endogenous active retinoids involved. The pleiotropic effects of retinoids may be explained by the existence of complex signal transduction pathways involving diverse nuclear receptors of the retinoic acid receptor (RAR) and retinoid X receptor (RXR) families, and at least two types of cellular retinoic acid binding proteins (CRABP-I and -II)². The different RARs, RXRs, and CRABPs have different expression patterns during vertebrate embryogenesis^{2,3}, suggesting that they each have particular functions. Another level at which fine tuning of retinoid action could occur is the metabolism of vitamin A to active metabolites,

which may include all-*trans*-retinoic acid^{4–7}, all-*trans*-3,4-didehydroretinoic acid⁸, 9-*cis*-retinoic acid^{9,10}, and 14-hydroxy-4,14-retroretinol¹¹. Formation of the metabolite all-*trans*-4-oxo-retinoic acid from retinoic acid was considered to be an inactivation pathway during growth and differentiation^{12–14}. We report here that, in contrast, 4-oxo-retinoic acid is a highly active metabolite which can modulate positional specification in early embryos. We also show that this retinoid binds avidly to and activates RAR β , and that it is available in early embryos. The different activities of 4-oxo-retinoic acid and retinoic acid in modulating positional specification on the one hand, and growth and differentiation on the other, interest us in the possibility that specific retinoid ligands regulate different physiological processes *in vivo*.

Treating *Xenopus laevis* embryos with retinoic acid (RA) during gastrulation causes a dose-dependent loss of anterior structures, leading to microcephaly (Fig. 1a, b; refs 6, 15, 16). We found that also 4-oxo-retinoic acid (4-oxo-RA) (see Fig. 1e for structural formula) is very active in causing microcephaly in *Xenopus* embryos, much more, in fact, than RA (Fig. 1a, b). The effects of 4-oxo-RA and RA on morphology appear qualitatively indistinguishable. The most sensitive stages for 4-oxo-RA treatment are the gastrula till mid-neurula stages (when the antero-posterior axis is specified), the same sensitive period reported previously for RA⁶ (not shown). Next, we tested whether 4-oxo-RA can induce the expression of Hox genes (class I homeobox containing genes), which are believed to confer anteroposterior positional information during vertebrate embryogenesis^{17–19}, and are inducible by RA^{18,19}. *Xenopus* embryos were treated with retinoids as in Fig. 1a, and expression of two *Xenopus* Hoxb genes, Hoxb-4 and Hoxb-9, was measured at several develop-

mental stages by RNase protection (Fig. 1c, d). Treatment with 10^{-6} M 4-oxo-RA caused strong induction of the expression of both Hoxb-4 and of Hoxb-9. 4-Oxo-RA, however induced the expression of both genes to roughly 2–3-fold higher levels compared with RA.

These strong effects of 4-oxo-RA on anteroposterior axis specification raise the possibility that this retinoid is a physiologi-

cal signal involved in providing the *Xenopus* embryo with antero-posterior positional information. A prerequisite for this possibility is that 4-oxo-RA be present endogenously in *Xenopus* embryos at the stages sensitive to treatment. We used high-performance liquid chromatography (HPLC) analysis to measure naturally occurring retinoids in *Xenopus* embryonic tissue. Retinoids were extracted from stage 10–13 embryos, and

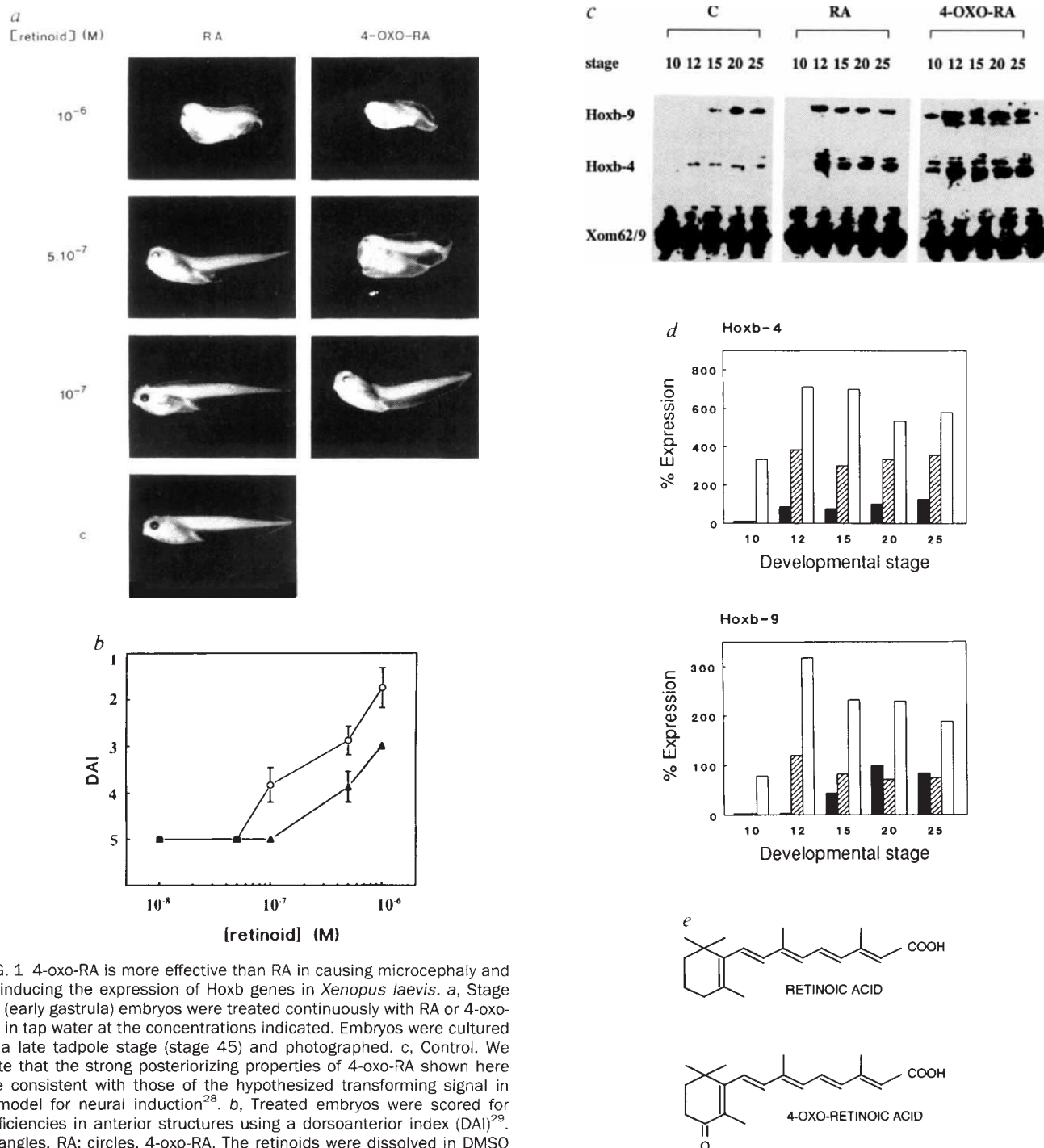


FIG. 1 4-oxo-RA is more effective than RA in causing microcephaly and in inducing the expression of Hoxb genes in *Xenopus laevis*. a, Stage 10 (early gastrula) embryos were treated continuously with RA or 4-oxo-RA in tap water at the concentrations indicated. Embryos were cultured to a late tadpole stage (stage 45) and photographed. c, Control. We note that the strong posteriorizing properties of 4-oxo-RA shown here are consistent with those of the hypothesized transforming signal in a model for neural induction²⁸. b, Treated embryos were scored for deficiencies in anterior structures using a dorsoanterior index (DAI)²⁹. Triangles, RA; circles, 4-oxo-RA. The retinoids were dissolved in DMSO and diluted with tap water immediately before use. Purity was checked by HPLC. Values are means $\pm$ s.e.m. ($n=10$). The experiment was repeated four times with similar results. c, RNase protection analysis of *Xenopus* Hoxb-4 and Hoxb-9 gene expression. Stage 10 embryos were treated with RA or 4-oxo-RA at 10^{-6} M as above. Embryos were frozen in liquid nitrogen at the developmental stages indicated, and processed for RNase protection analysis as described¹⁹. For RNase protection of stage 10 embryos, embryos were treated with retinoids

for 30 min. Xom 62/9 was used as an internal standard. The experiment was performed three times with similar results. c, Control. d, Quantification of the expression data from c by phosphorimaging (Molecular Dynamics). Black bars, no treatment; hatched bars, RA; empty bars, 4-oxo-RA. The data are expressed as percentages relative to the expression in untreated embryos at stage 20. e, Structural formulas of RA and 4-oxo-RA.

free acids were methylated using diazomethane²⁰. The extract was then separated by reverse-phase HPLC. We observed a peak which coeluted exactly with standard methylated 4-oxo-RA (4-oxo-methyl retinoate: Me-4-oxo-RA; Fig. 2a). This methylated compound now eluted separately from non-retinoid impurities which elute at the position of 4-oxo-RA during HPLC analysis of non-methylated samples. The endogenous peak coeluting with Me-4-oxo-RA on reverse-phase HPLC was collected and analysed by normal-phase HPLC (Fig. 2b). The resulting chromatogram contained two peaks, one of which appeared to be a non-

retinoid compound (by ultraviolet-spectrum analysis, not shown), and one of which coeluted exactly with standard Me-4-oxo-RA. We next used photodiode array detection to measure the absorption spectrum of the putative endogenous Me-4-oxo-RA. Figure 2c shows an overlay of the absorption spectra of standard and of putative endogenous Me-4-oxo-RA, recorded under the same conditions. The spectra of the two compounds were identical, with maximal absorption at 354 nm. These data indicate that 4-oxo-RA is present in the *Xenopus* embryo during anteroposterior axis specification. We measured a mean concen-

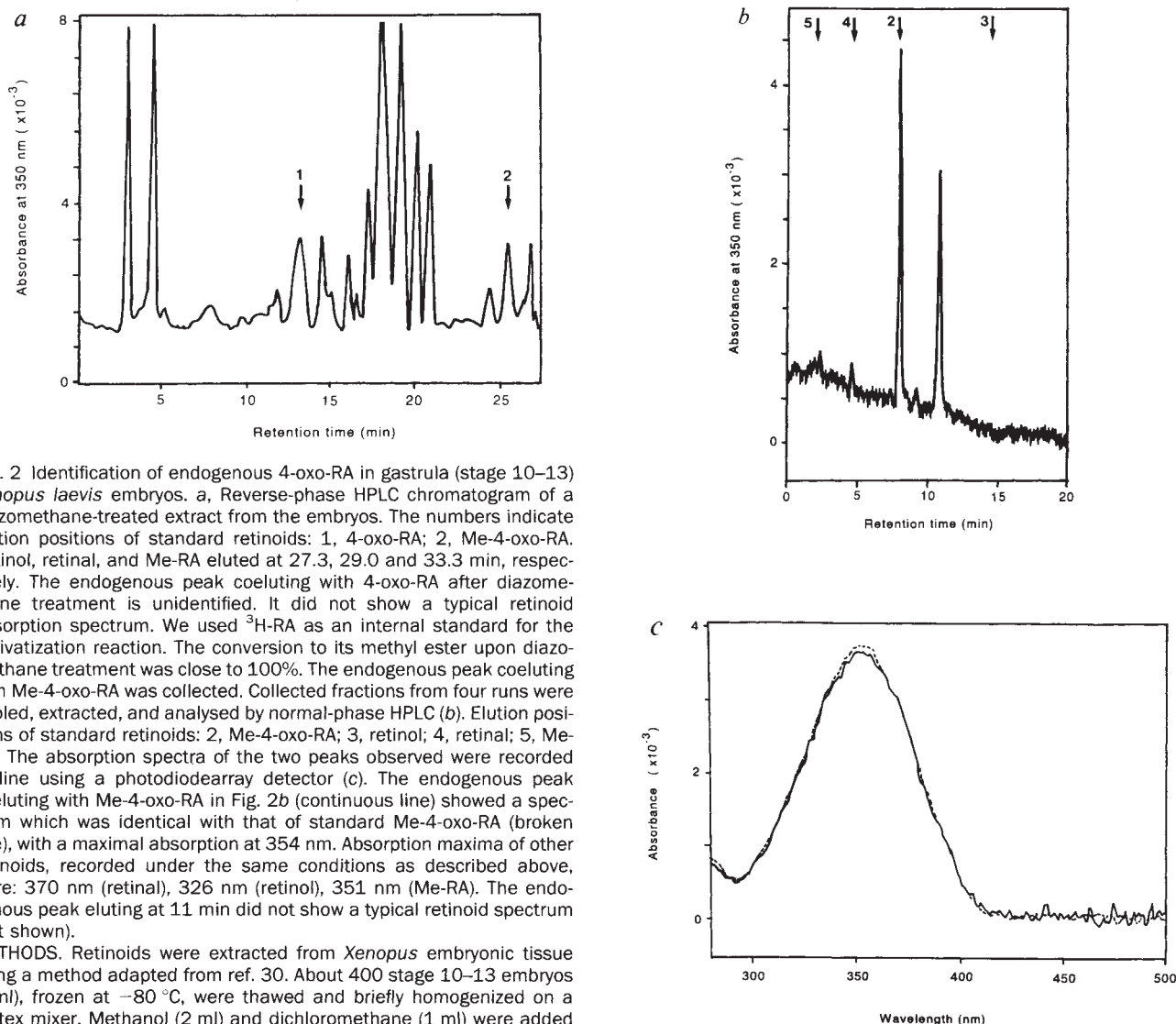


FIG. 2 Identification of endogenous 4-oxo-RA in gastrula (stage 10–13) *Xenopus laevis* embryos. a, Reverse-phase HPLC chromatogram of a diazomethane-treated extract from the embryos. The numbers indicate elution positions of standard retinoids: 1, 4-oxo-RA; 2, Me-4-oxo-RA. Retinol, retinal, and Me-RA eluted at 27.3, 29.0 and 33.3 min, respectively. The endogenous peak coeluting with 4-oxo-RA after diazomethane treatment is unidentified. It did not show a typical retinoid absorption spectrum. We used ³H-RA as an internal standard for the derivatization reaction. The conversion to its methyl ester upon diazomethane treatment was close to 100%. The endogenous peak coeluting with Me-4-oxo-RA was collected. Collected fractions from four runs were pooled, extracted, and analysed by normal-phase HPLC (b). Elution positions of standard retinoids: 2, Me-4-oxo-RA; 3, retinol; 4, retinal; 5, Me-RA. The absorption spectra of the two peaks observed were recorded on-line using a photodiodearray detector (c). The endogenous peak coeluting with Me-4-oxo-RA in Fig. 2b (continuous line) showed a spectrum which was identical with that of standard Me-4-oxo-RA (broken line), with a maximal absorption at 354 nm. Absorption maxima of other retinoids, recorded under the same conditions as described above, were: 370 nm (retinal), 326 nm (retinol), 351 nm (Me-RA). The endogenous peak eluting at 11 min did not show a typical retinoid spectrum (not shown).

METHODS. Retinoids were extracted from *Xenopus* embryonic tissue using a method adapted from ref. 30. About 400 stage 10–13 embryos (1 ml), frozen at -80°C , were thawed and briefly homogenized on a Vortex mixer. Methanol (2 ml) and dichloromethane (1 ml) were added and the mixture was vortexed for 4 min. The residue was removed by filtration through a glass sinter. The residue was dissolved in 2 ml methanol and 1 ml dichloromethane, and 5 ml dichloromethane was added. This was vortexed for 4 min, and filtered as above. The two filtrates were combined, 1.75 ml 0.9% NaCl was added, and the mixture was vortexed for 4 min. The mixture was allowed to settle, and the upper layer was removed. The bottom layer was evaporated using a stream of nitrogen. The dry lipids were dissolved by addition of 100 μl 60 mM ammonium acetate, pH 5.75, followed by 300 μl methanol. This mixture was transferred to a small glass tube and centrifuged for 15 s at 1,000g. The upper layer was treated for 5 min on ice with an ethereal solution of diazomethane²⁰. The solvent was then evaporated with a stream of nitrogen and the lipids were dissolved in 125 μl methanol. The extract was injected into a reverse-phase HPLC system containing a Spherisorb (Phase Separations) S50DS2 guard column (5×0.46 cm) and an identical analytical column (25×0.46 cm). Retinoids were detected by measuring the absorbance at 350 nm. Extracts were separated using gradient elution with solvent A²⁰ (60 mM ammonium acetate,

pH 5.75 (adjusted with methanol:acetic acid, 1:1)) and solvent B (methanol). The gradient program with a flow rate of 1 ml min^{-1} was: 5 min isocratic at 65% B, followed by a convex gradient (Waters, number 4) to 85% in 15 min, and a further convex gradient (as above) to 99% B in 15 min. Collected fractions were extracted twice (to remove mobile phase salts) by partition between water (0.9 volume) and dichloromethane (1 volume). The organic phase was evaporated to complete dryness, and the residue was dissolved in 125 μl hexane/dioxane (25/1), which was the mobile phase used for isocratic normal-phase HPLC. The normal-phase HPLC system contained a Spherisorb (Phase Separations) S3W silica column (10×0.46 cm) linked to a Waters model 994 photodiodearray detector, which was used for recording the chromatogram at 350 nm as shown in Fig. 2b, and for spectral analysis as shown in Fig. 2c. Retinoids were quantified using external (4-oxo-RA) and internal (³H-RA) standards. All solvents used were HPLC grade. All handling of retinoids and biological samples was done in a nitrogen atmosphere in a dark room illuminated with red light.

tration of 20 nM endogenous 4-oxo-RA at the gastrula stages (stages 10–13).

In analogy with RA^{2,21}, all-*trans*-3,4-didehydroretinoic acid (ddRA)^{2,21}, and 9-*cis*-retinoic acid (9-*cis*-RA)^{2,9,10,21}, the biological effects of 4-oxo-RA are probably mediated by nuclear hormone receptors, possibly members of the RAR or RXR families. We therefore investigated whether 4-oxo-RA can activate two representative members of these families: RAR β and RXR α . To avoid interference by endogenous receptors, we used chimaeric receptors consisting of domains D–F of RAR β or domains C–E of RXR α , fused to the DNA binding domain of GAL4 (GAL-DBD)^{21,23}. Figure 3a shows that 4-oxo-RA is a very potent activator of GAL4-RAR β , and that the activity of 4-oxo-RA is similar to that of RA and 9-*cis*-RA. In contrast, 4-oxo-RA is a weak activator of GAL4-RXR α resembling RA in this respect (Fig. 3b). As expected, 9-*cis*-RA activated GAL4-RXR α efficiently. To test whether 4-oxo-RA actually binds to RAR β , we did a ligand binding assay using nuclear extracts from Cos-1 cells transfected with an expression plasmid containing full-length RAR β complementary DNA^{9,21}. Figure 3c shows the results of a competition binding assay in which unlabelled RA or 4-oxo-RA were tested for their ability to compete with a fixed amount of ³H-RA for binding to RAR β . The results show that 4-oxo-RA binds RAR β with high affinity, similar to RA, which is consistent with a previous report about the binding of retinoids to skin RARs²⁴. Together with the transactivation data these

data indicate that 4-oxo-RA is a high-affinity, activating ligand for at least one RAR, namely RAR β .

Besides its effects on embryogenesis, RA also has strong effects on growth and differentiation in various cell lines and tissues. The activity of 4-oxo-RA is much lower than that of RA in a number of growth and differentiation assays^{12–14}. We extended these observations by analysing the effects of RA and 4-oxo-RA on the growth and differentiation of murine embryonic stem (ES) cells. RA stimulates ES cell growth at low concentrations (10⁻¹⁰–10⁻⁹ M) whereas it inhibits growth at higher concentrations (10⁻⁸–10⁻⁶ M) (Fig. 4). The effects of 4-oxo-RA on growth qualitatively resemble those of RA, but its activity is lower than that of RA, both at high and low concentrations. Maximal growth stimulation by low 4-oxo-RA concentrations is 30%, whereas RA is able to stimulate growth by 75%. At higher concentrations, the same level of growth inhibition is achieved with a roughly 10-fold higher dose of 4-oxo-RA as compared with RA. The observed growth inhibition at higher retinoid concentrations was accompanied by an induction of differentiation, primarily to parietal endoderm cells. In this case, also, the activity of RA was clearly higher than that of 4-oxo-RA (not shown).

Our findings that 4-oxo-RA is highly active in modifying positional specification and is endogenously present in the early embryo suggest a role for 4-oxo-RA as a physiological retinoid signal involved in anteroposterior patterning. We note, that the activity of 4-oxo-RA is higher than that of RA, both in causing

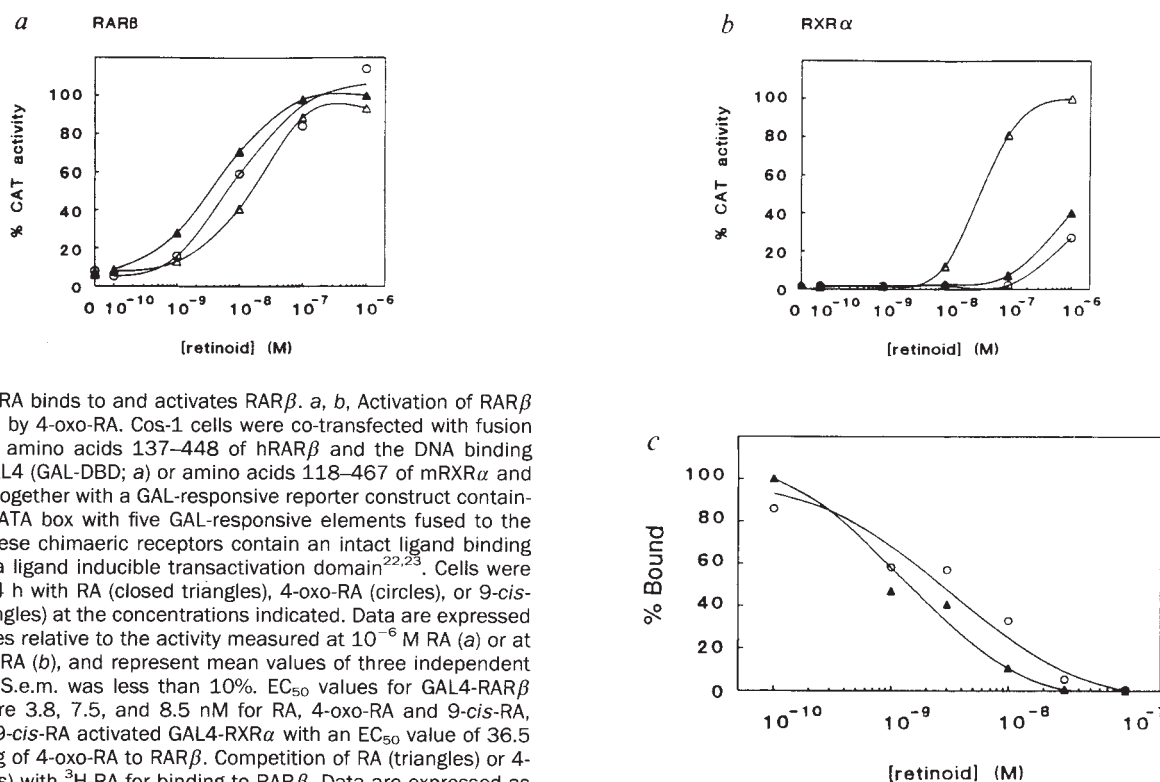


FIG. 3 4-oxo-RA binds to and activates RAR β . *a, b*, Activation of RAR β but not RXR α by 4-oxo-RA. Cos-1 cells were co-transfected with fusion constructs of amino acids 137–448 of hRAR β and the DNA binding domain of GAL4 (GAL-DBD; *a*) or amino acids 118–467 of mRXR α and GAL-DBD (*b*) together with a GAL-responsive reporter construct containing an E1b TATA box with five GAL-responsive elements fused to the CAT gene. These chimaeric receptors contain an intact ligand binding domain and a ligand inducible transactivation domain^{22,23}. Cells were treated for 24 h with RA (closed triangles), 4-oxo-RA (circles), or 9-*cis*-RA (open triangles) at the concentrations indicated. Data are expressed as percentages relative to the activity measured at 10⁻⁶ M RA (*a*) or at 10⁻⁶ M 9-*cis*-RA (*b*), and represent mean values of three independent experiments. S.e.m. was less than 10%. EC₅₀ values for GAL4-RAR β activation were 3.8, 7.5, and 8.5 nM for RA, 4-oxo-RA and 9-*cis*-RA, respectively; 9-*cis*-RA activated GAL4-RXR α with an EC₅₀ value of 36.5 nM. *c*, Binding of 4-oxo-RA to RAR β . Competition of RA (triangles) or 4-oxo-RA (circles) with ³H-RA for binding to RAR β . Data are expressed as percentages specific binding. IC₅₀ values were 0.9 and 4.2 nM for RA and 4-oxo-RA, respectively.

METHODS. The construction of the GAL4-RAR β expression plasmid has been described²³. The GAL4-RXR α expression plasmid was constructed by the cloning of the *Stu1*–*Sma1* fragment from mRXR α into the *Sma1* site of pSG424 (ref. 23). pSG5-RXR α was provided by P. Chambon. Cos-1 cells were cultured and transfected using calcium phosphate as described²³. Cells were co-transfected with SV2-lacZ as an internal control for transfection efficiency. For transactivation studies, precipitates were removed 18 h after transfection, and retinoids were added in fresh medium. After an additional 24 h period cells were assayed for CAT activity as described²³. Results were quantified using a phosphorimager (Molecular Dynamics). EC₅₀ values were calculated assuming a sigmoidal dose response with a 100% maximal value. The ligand binding

assay was done essentially as described previously^{9,21}. In short, Cos-1 cells were transfected as above with a pSG5 expression plasmid containing the full-length hRAR β cDNA²³. 48 h after transfection, nuclear extracts were prepared^{9,21} and stored at –80 °C until use. Competition binding assays were done by addition of aliquots of RAR β nuclear extracts to 0.2 nM ³H-RA (NEN, 55.7 Ci mmol⁻¹) mixed with various concentrations of unlabelled competitor retinoid. The incubation volume was 100 μ l. After incubation in the dark at 4 °C for 4 h, bound ligand was separated from free using Pharmacia PD10 columns as described^{9,21}. Fractions were counted using a liquid scintillation counter. Nuclear extracts from cells transfected with pSG5 without the RAR β cDNA did not show specific binding of ³H-RA. IC₅₀ values were calculated as above.

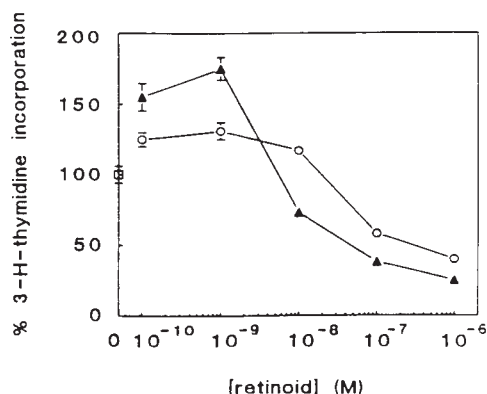


FIG. 4 Effects of RA (triangles) and 4-oxo-RA (circles) on growth of ES cells.

METHODS. ES cells (E14 (ref. 31)) were plated in 24-well plates at 10,000 cells per well in 80% Buffalo rat liver cell-conditioned medium (BRL-CM)/20% minimal essential medium (MEM), containing 20% fetal calf serum. After plating, retinoids were added to yield the concentrations indicated. 66 h after plating, ³H-thymidine (Amersham, UK) was added at 0.5 μ Ci per well. After an additional 8 h period cells were washed with PBS, fixed with methanol for 15 min, air dried, and incubated with 0.1 M NaOH for 30 min at 37 °C. NaOH extracts were counted in a liquid scintillation counter. Data are expressed as percentages relative to control incorporation (no treatment; boxes) and represent mean values $\pm$ s.e.m. ($n = 4$).

microcephaly and in inducing expression of Hoxb genes, and further that 4-oxo-RA is a major RA metabolite both *in vitro*¹² and *in vivo*²⁵ which cannot be converted back to RA²⁶. *Xenopus* gastrula embryos also metabolize exogenously applied RA to a number of products which include 4-oxo-RA (data not shown). The effects of RA on anteroposterior axis specification might thus occur at least partially by metabolism to 4-oxo-RA. But we cannot exclude the possibility that RA itself also acts directly as an active ligand during this process, perhaps by mimicking the action of 4-oxo-RA, because these two retinoids share common receptors and binding proteins. These include RAR β , and presumably also CRABP²⁷.

In contrast to the high activity of 4-oxo-RA during anteroposterior axis formation, this retinoid is less active than RA in a number of growth and differentiation assays, including retinoid modulated growth and differentiation of ES cells. Apparently, the relative activities of RA and 4-oxo-RA depend on the retinoid-response system. This raises the possibility that these different activities reflect different *in vivo* functions of RA and 4-oxo-RA. 4-oxo-RA might have a specific function in providing positional information, whereas RA itself or another RA metabolite has a specific function in controlling growth and differentiation. Further research should determine the validity of this hypothesis, and elucidate the functions of these different retinoid ligands during growth, differentiation, and anteroposterior axis formation. □

Received 11 February; accepted 17 September 1993.

- Maden, M. & Holder, H. *Bioessays* **14**, 431–438 (1992).
- Leid, M., Kastner, P. & Chambon, P. *Trends Biochem. Sci.* **17**, 427–433 (1992).
- Ellinger-Ziegelbauer, H. & Dreyer, C. *Genes Dev.* **5**, 94–104 (1991).
- Tickle, C., Alberts, B., Wolpert, L. & Lee, J. *Nature* **296**, 564–566 (1982).
- Thaller, C. & Eichele, G. *Nature* **327**, 625–628 (1987).
- Durston, A. J. *et al. Nature* **340**, 140–144 (1989).
- Hogan, B. L. M., Thaller, C. & Eichele, G. *Nature* **359**, 237–241 (1992).
- Thaller, C. & Eichele, G. *Nature* **345**, 815–819 (1990).
- Levin, A. A. *et al. Nature* **355**, 359–361 (1992).
- Heyman, R. A. *et al. Cell* **68**, 397–406 (1992).
- Buck, J. *et al. Science* **254**, 1654–1656 (1991).
- Frolik, C. A. *et al. Biochemistry* **18**, 2092–2097 (1979).
- Surekha Rao, M. S., John, J. & Cama, H. R. *Int. J. Vitam. Nutr. Res.* **42**, 368–378 (1972).
- Williams, J. B., Shoelds, C. O., Brettell, L. M. & Napoli, J. L. *Analyt. Biochem.* **160**, 267–274 (1987).
- Sive, H. L., Draper, B. W., Harland, R. M. & Weintraub, H. *Genes Dev.* **4**, 932–942 (1990).

- Ruiz i Altaba, A. & Jessell, T. *Genes Dev.* **5**, 175–187 (1991).
- Wilkinson, D. G. *et al. Nature* **341**, 405–409 (1989).
- Simeone, A. *et al. Nature* **346**, 763–766 (1990).
- Dekker, E. J. *et al. Mech. Dev.* **40**, 3–12 (1992).
- Eckhoff, C. & Nau, H. J. *Lipid Res.* **31**, 1445–1454 (1990).
- Allenby, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 30–34 (1993).
- Nagpal, S. *et al. Cell* **70**, 1007–1019 (1992).
- Folkers, G. E., van der Leede, B. M. & van der Saag, P. T. *Molec. Endocr.* **7**, 616–627 (1993).
- Sani, B. P. in *Retinoids* (eds Livrea, M. A. & Packer, L.) 237–247 (Marcel Dekker, New York, 1993).
- Frolik, C. A., Roller, P. P., Roberts, A. B. & Sporn, M. B. *J. biol. Chem.* **255**, 8057–8062 (1980).
- Roberts, A. B., Lamb, L. C. & Sporn, M. B. *Arch. Biochem. Biophys.* **199**, 374–383 (1980).
- Fiorella, P. D. & Napoli, J. L. *J. biol. Chem.* **266**, 16572–16579 (1991).
- Nieuwkoop, P. D. *et al. J. exp. Zool.* **120**, 1–108 (1952).
- Kao, K. R. & Elinson, R. P. *Dev. Biol.* **127**, 64–77 (1988).
- Palmer, F. B. St. C. *Biochim. biophys. Acta* **231**, 134–144 (1971).
- Handyside, A. H., O'Neill, G. T., Jones, M. & Hooper, M. L. *Roux's Arch. dev. Biol.* **198**, 48–55 (1989).

ACKNOWLEDGEMENTS. We thank J. Bausch (Hoffmann-La Roche, Basel) for his gift of retinoids, W. E. Lambert and A. N. Sakke for advice on the HPLC technique, A. Levin and J. Boonstra for advice on the ligand binding assay, P. Chambon for providing pSG5-RXR α and pSG5-RAR β , S. W. de Laat for discussion, M. Peppelenbosch for assistance with data analysis and A. A. Glaudemans, F. Vervoordeldonk and J. Heinen for preparing the figures. This work was supported by the Netherlands Organisation for Scientific Research Council for Medical Research (W. P.), Council for Biological Research (E. D.) and the Netherlands Cancer Foundation (G. F.).

Dopamine D1 receptors facilitate transmitter release

Donald L. Cameron & John T. Williams*

Vollum Institute for Advanced Biomedical Research,
Oregon Health Sciences University, Portland, Oregon 97201, USA

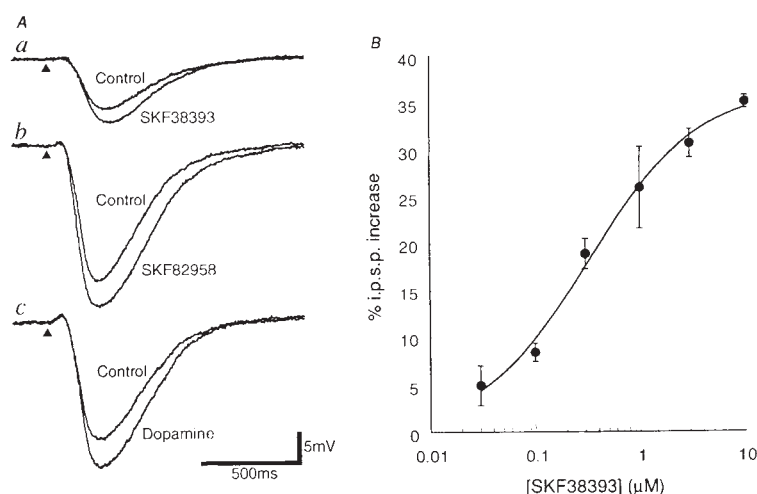
A PHYSIOLOGICAL role for the dopamine D1 receptor has been difficult to define, particularly because of its complex pre- and postsynaptic localization in brain areas such as the striatum¹. In the midbrain, however, D1 receptors are selectively localized to the terminals of GABA (γ -aminobutyric acid)-containing afferents^{2–4}. We have studied the actions of these D1 receptors on evoked GABA synaptic potentials recorded intracellularly from dopamine neurons in the ventral tegmental area (VTA). We report here that dopamine augmented GABA_B inhibitory postsynaptic potentials (i.p.s.ps) in the presence of D2 receptor antagonists. This effect was mimicked by the D1 agonists SKF38393 and SKF82958 and blocked by the D1 antagonists SCH23390 and *cis*-flupenthixol. No modulation of the GABA_A synaptic potential was observed. The postsynaptic actions of the GABA_B agonist, baclofen, were unaffected by SKF38393, SCH23390 or *cis*-flupenthixol, confirming a presynaptic locus of D1 action. Additionally, D1 antagonists reduced the amplitude of the GABA_B i.p.s.p. in the absence of D1 agonists. We conclude that dopamine acts tonically at presynaptic D1 receptors on the terminals of afferent GABA neurons to facilitate selectively GABA_B-mediated neurotransmission in the midbrain.

Evoked i.p.s.ps were recorded from dopamine neurons in horizontal slices of guinea-pig VTA. Dopamine neurons were identified by their characteristic membrane properties and their hyperpolarizing response to dopamine⁵. Fast GABA_A-, 2-(amino-methyl)phenylacetic acid (AMPA)- and NMDA (*N*-methyl-D-aspartate)-mediated synaptic potentials were blocked by the addition of picrotoxin, 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and 2-amino-5-phosphonopentanoic acid (AP5), respectively. Recorded i.p.s.ps were confirmed to be mediated by GABA_B receptors by antagonism with either saclofen (100 μ M) or CGP35348 (100 μ M). In the presence of the selective D2 antagonists eticlopride or sulpiride, dopamine (10 μ M) augmented the amplitude of the i.p.s.p. (8 out of 12 neurons tested; Fig. 1a). These responses varied widely and may have

* To whom correspondence should be addressed.

FIG. 1 Superimposed membrane potential traces showing facilitation of the GABA_B i.p.s.p. by Aa, SKF38393 (1 μ M); Ab, SKF82958 (1 μ M); and Ac, dopamine (10 μ M). Recordings were made in current clamp with the resting membrane potential maintained between -65 and -70 mV. Each trace is an average of 4–5 i.p.s.ps with the beginning of the train of stimuli used to evoke the i.p.s.p. indicated by the triangle. B, concentration-response curve for SKF38393 showing the per cent increase in peak i.p.s.p. amplitude compared to control. Each point is the mean $\pm$ s.e.m. of 3–4 determinations at each concentration SKF38393. Measurements of amplitude changes were only made after the effect of SKF38393 had reached steady-state, generally 6–10 min.

METHODS. Intracellular recordings of membrane potential were made from neurons in horizontal slices of guinea-pig midbrain. This method has been described previously²⁹. Briefly, guinea pigs (300–400 g) were anaesthetized with halothane and killed. The brain was subsequently removed and the midbrain sliced (300 μ M) in the horizontal plane using a Vibratome. The slice containing the VTA was placed in a recording chamber and superfused (1.5 ml min⁻¹) with warmed (35 $^{\circ}$ C) Krebs/bicarbonate buffer containing the following (in mM): NaCl 126, KCl 2.5, NaH₂PO₄ 1.2, MgCl₂ 1.2, CaCl₂ 2.4, glucose 11, NaHCO₃ 21.4 saturated with 95% O₂ and 5% CO₂. The VTA was taken as the area lying between, and rostral to, the interpeduncular fossa (IPF) and the medial terminal nucleus accessory optic tract (MT). Bipolar tungsten stimulating electrodes were placed medio-caudal and rostral to the MT. To evoke synaptic potentials, either a single stimulus or trains of stimuli (500 μ s at 70 Hz for 143 ms, 10 stimuli) ranging from 0.5–1.5 mA were delivered at 90-s intervals using a constant-current stimulation unit. Recordings were made with 2M KCl-filled glass



microelectrodes (20–50 M Ω) using standard techniques. Drugs were applied in known concentrations to the superfusion medium. In experiments examining the GABA_B synaptic potential, the superfusion medium contained 100 μ M 2-amino-5-phosphonopentanoic acid (AP5), 10 μ M 6-cyano-2,3-dihydroxy-7-nitro-quinoline (CNQX) and 100 μ M picrotoxin to block fast NMDA, AMPA, and GABA_A mediated synaptic potentials, respectively. Picrotoxin was omitted in the experiments examining the GABA_A synaptic potential. In all experiments either 1 μ M sulpiride or 100 nM eticlopride was included in the superfusion solution to block any possible effect mediated by the dopamine D2 receptor.

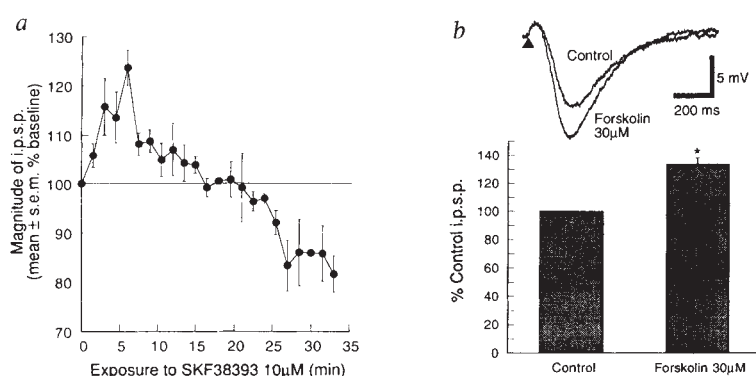
been complicated by the extensive dopamine re-uptake present in the slice and the presence of endogenous dopamine (see below). But this effect was mimicked (26 out of 28 neurons tested) by the selective D1 agonists SKF38393 and SKF82958 (Fig. 1a). No change in resting membrane potential or the time course of the i.p.s.p. was observed.

The amplitude of the GABA_B i.p.s.p. was increased in a concentration-dependent manner by SKF38393 (30 nM–10 μ M, Fig. 1b). The half maximal effect (EC₅₀) was achieved at a concentration of 327 nM which is similar to that obtained for the D1-mediated stimulation of cyclic AMP by SKF38393 in striatal homogenates⁶. The maximal increase in i.p.s.p. amplitude (E_{max}) was 36% and when a maximal effective concentration of SKF38393 (10 μ M) was applied, the facilitation effect was observed to peak after 5–7 min. Subsequently the amplitude of

the i.p.s.p. declined over a period of 30 min to about 20% below that of control (Fig. 2a). The time course of this 'run-down' parallels the agonist-induced desensitization observed for the D1 receptor mediated stimulation of adenylate cyclase in slices and cultures^{7,8}. These observations may point to a role for cAMP as an intracellular messenger in the facilitation effect and indeed, direct stimulation of adenylate cyclase by bath-applied forskolin (30 μ M, Fig. 2b) mimicked the action of the D1 agonists. Studies on isolated chromaffin cells⁹ have indicated that D1 receptors may enhance calcium currents through a cAMP-dependent process, suggesting a potential mechanism whereby transmitter release might be enhanced.

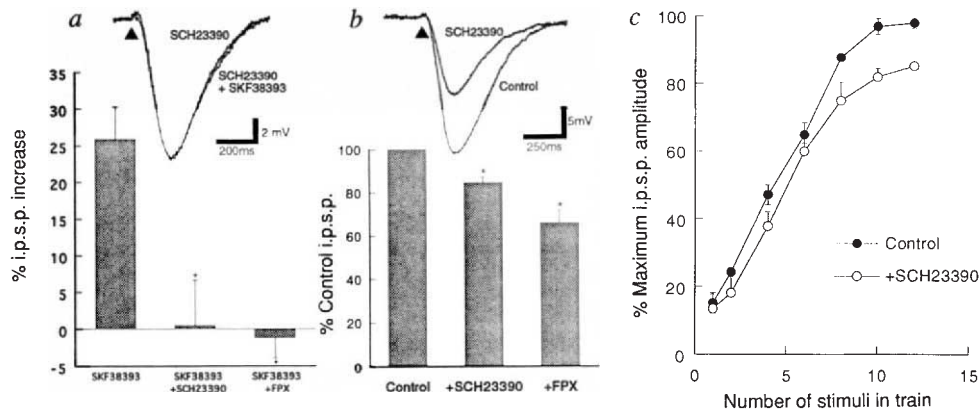
The observation that the 'run-down' of the D1 effect ultimately reached a level below that of control implies that the D1 receptors in the slice are under tonic, sub-maximal stimulation.

FIG. 2 a, Time course of change in the amplitude of the GABA_B i.p.s.p. in a high concentration of D1 agonist. The change in amplitude to SKF38393 (10 μ M) was assessed at 1.5 min intervals, with the maximal response occurring between 5 and 7 min after the beginning of the exposure to the agonist. After this point the response declined until it reached control levels at 20 min after the start of agonist application. The amplitude of the i.p.s.p. continued to decrease until it stabilized at about 80% of control amplitude some 30 min after the start of the agonist application. The graph points represent the mean $\pm$ s.e.m. of 4 experiments. b, The effect of adenylate cyclase stimulation on the amplitude of the GABA_B i.p.s.p. When forskolin (30 μ M) was applied in the superfusion medium, the magnitude of the i.p.s.p. increased significantly (136.7 $\pm$ 4.2% of control amplitude, P < 0.05, Wilcoxon test, n = 5). The maximal forskolin effect occurred after a period of about 10 min. Inset, superimposition of GABA_B i.p.s.ps in the presence and absence of forskolin (30 μ M). Each trace in the average of 4 i.p.s.ps.



The beginning of the train of stimuli used to evoke the i.p.s.p. is indicated by the triangle.

FIG. 3 a, The effect of D1 antagonists on the facilitation of the GABA_B i.p.s.p. by SKF38393. When cells were exposed to SCH23390 (10 nM) or *cis*-flupenthixol (FPX, 10 μ M), the effects of SKF38393 (1 μ M) were completely blocked ($P < 0.05$, Kruskal-Wallis test, $n = 3$ per group). Inset, superimposition of GABA_B i.p.s.ps with and without SKF38393 (1 μ M) in the presence of SCH23390 (10 nM). Each trace is the average of 4–5 i.p.s.ps. The beginning of the train of stimuli used to evoke the i.p.s.p. is indicated by the triangle. b, D1 antagonists reduced the amplitude of the GABA_B i.p.s.p. under control conditions. Top trace (average of 5 i.p.s.ps) shows an example of the effect of SCH23390 (1 μ M) on the amplitude of the i.p.s.p. The beginning of the train of stimuli used to evoke the i.p.s.p. is indicated by the triangle. In separate experiments, both SCH23390 (1 μ M) and *cis*-flupenthixol (10 μ M) were observed to reduce the GABA_B i.p.s.p. to $84.6 \pm 2.6\%$ and $65 \pm 5.9\%$ of the control amplitude, respectively ($P < 0.05$ Kruskal-Wallis test, $n =$



3–4 for each group). c, Stimulus–response curve for the GABA_B i.p.s.p. in the presence and absence of SCH23390 (1 μ M). All stimuli were delivered at 70 Hz. The D1 antagonist SCH23390 significantly reduced the amplitude of the i.p.s.p. with train lengths consisting of 8 or more stimuli. Points are mean $\pm$ s.e.m. of 3–4 determinations at each stimulus value.

While the D1 antagonists SCH23390 or *cis*-flupenthixol blocked the agonist-induced augmentation of the i.p.s.p. (Fig. 3a), both antagonists caused a reduction in the amplitude of the i.p.s.p. when applied alone (Fig. 3b). This effect was dependent on the length of the stimulus train (Fig. 3c), with the longer train lengths presumably causing an additional release of endogenous dopamine. Taken together, these data suggest that stimulation of D1 receptors by endogenous dopamine, released from the dendrites of spontaneously active VTA dopamine neurons^{10,11}, results in a tonic enhancement of GABA release.

Evidence supporting the presynaptic nature of the D1 effect was gained from two sets of observations. First, SKF38393, SCH23390 or *cis*-flupenthixol did not affect the hyperpolarizing action of bath-applied baclofen (Fig. 4a). Second, SKF38393 (1 μ M) caused a significant increase in the ratio of the amplitudes of two i.p.s.ps produced by paired trains of stimuli 750 ms apart (0.71 versus 0.76, $n = 5$, $P < 0.05$ Wilcoxon signed rank test). This is consistent with a presynaptic locus of action for the D1 agonist¹².

SKF38393 failed to facilitate evoked GABA_A synaptic potentials (Fig. 4b), demonstrating that the D1 effect was selective for the GABA_B i.p.s.p. As the D1 receptors in the VTA are localized on the terminals of GABAergic afferents arising from the forebrain^{2–4}, this finding supports the proposal that GABA_B receptors are selectively present at synapses formed by forebrain GABAergic afferents, whereas GABA_A receptors are localized at synapses formed with intrinsic GABA interneurons^{13,14}.

Thus, the action of dopamine D1 receptors in enhancing GABA release in the midbrain is both anatomically and receptor specific, modulating the 'slow' synaptic actions of GABA as part of the feedback circuit from the forebrain. A functional consequence of this arrangement is that increased midbrain dopamine release will increase the efficacy of any subsequent activity in the descending GABA afferents. This effect will be self-limiting as the inhibitory action of GABA on the dopamine neurons should lead to a subsequent reduction in midbrain dopamine release.

Many transmitters act presynaptically to modulate synaptic release¹⁵ and examples of presynaptic facilitation have been found in the mammalian peripheral nervous system¹⁶, *Aplysia* ganglia^{17,18} and in teleost Mauthner cells¹⁹. But all of the presynaptic actions demonstrated electrophysiologically in the central nervous system have been inhibitory in nature. Thus, to our knowledge, this action of D1 receptors is the first example in

the mammalian central nervous system of a presynaptic receptor directly facilitating, rather than inhibiting, the release of another neurotransmitter. Previous work using extracellular recordings²⁰ or evoked [³H]GABA release^{21,22} has supported a facilitating role for midbrain D1 receptors in GABA-mediated neurotransmission. But these studies have been unable to distinguish between direct and indirect D1 effects.

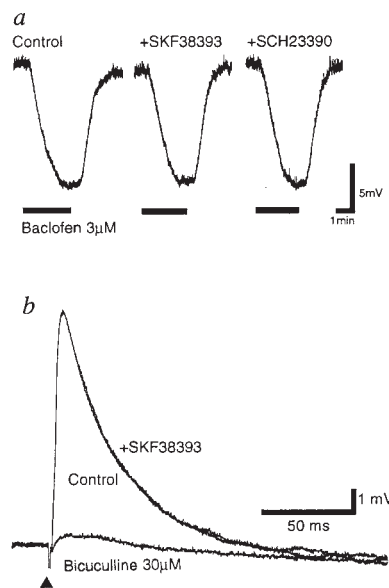


FIG. 4 a, The effect of D1 receptor agonist and antagonists on the amplitude of membrane hyperpolarization induced by the GABA_B agonist, baclofen. Baclofen (3 μ M) was bath applied and the resulting membrane hyperpolarization was assessed in the presence and absence of SKF38393 (1 μ M), SCH23390 (1 μ M) and Flupenthixol (10 μ M). None of these drugs produced any difference in the maximal effect of baclofen ($n = 3–5$ cells per group). b, The GABA_A synaptic potential was not affected by SKF38393 (1 μ M) but was abolished by the GABA_A antagonist bicuculline (30 μ M). The GABA_A synaptic potential (each trace is the average of 4 synaptic potentials) was evoked by a single stimulus of 500 μ s duration in the absence of picrotoxin in the perfusion medium ($n = 3$ cells per group).

A further consideration is that, although five dopamine receptor subtypes have been identified by molecular cloning techniques, only D1-like (D1, D5) and D2-like (D2, D3, D4) receptors can be readily distinguished pharmacologically²³. Consequently, the potential involvement of the D5 receptor cannot be excluded.

The D1-mediated facilitation of GABA transmission at GABA_B receptors suggests a simple cellular model for the D1 facilitation of D2-mediated responses observed in the whole animal²⁴⁻²⁶. Dopamine acting at D2 receptors hyperpolarizes mid-brain dopamine neurons by increasing potassium conductance²⁷.

Similarly, GABA acting at GABA_B receptors can act by the same mechanism in the same cells²⁸. Thus, the costimulation of D1 and D2 receptors may result in a larger increase in cellular potassium conductance than that afforded by stimulation of the D2 receptor alone.

Defining of the role of D1 receptors in the midbrain may also have important implications for understanding movement disorders, such as Parkinson's disease and the motor side-effects caused by dopamine receptor antagonists in the form of antipsychotic drugs. □

Received 12 August; accepted 21 September 1993.

1. Surmeier, D. J., Reiner, A., Levine, M. S. & Ariano, M. A. *Trends Neurosci.* **16**, 299-305 (1993).
2. Harrison, M. B., Wiley, R. G. & Wooten, G. F. *Brain Res.* **528**, 317-322 (1990).
3. Mansour, A. et al. *Neuroscience* **42**, 359-371 (1991).
4. Filloux, F. M., Dawson, T. M. & Wamsley, J. K. *Brain Res. Bull.* **20**, 447-459 (1988).
5. Johnson, S. W. & North, R. A. *J. Physiol.* **450**, 455-468 (1992).
6. Mottola, D. M., Brewster, W. K., Cook, L. L., Nichols, D. E. & Mailman, R. B. *J. Pharmac. exp. Ther.* **262**, 383-393 (1992).
7. Memo, M., Lovenberg, W. & Hanbauer, I. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4456-4460 (1982).
8. Baimforth, A. J., Warburton, P. & Ball, S. G. *J. Neurochem.* **55**, 2111-2116 (1990).
9. Artalejo, C. R., Ariano, M. A., Perlman, R. L. & Fox, A. P. *Nature* **346**, 239-242 (1990).
10. Geffen, L. B., Jessell, T. M., Cuelllo, A. C. & Iversen, L. L. *Nature* **260**, 258-260 (1976).
11. Llinás, R., Greenfield, S. A. & Jahnsen, H. *Brain Res.* **294**, 127-132 (1984).
12. Davies, C. H., Davies, S. N. & Collingridge, G. L. *J. Physiol.* **424**, 513-531 (1990).
13. Sugita, S., Johnson, S. W. & North, R. A. *Neurosci. Lett.* **134**, 207-211 (1992).
14. Johnson, S. W., Mercuri, N. B. & North, R. A. *J. Neurosci.* **12**, 2000-2006 (1992).
15. Chesselet, M. F. *Neuroscience* **12**, 347-375 (1984).
16. Zizgias, J. & Cunnance, T. C. *Br. J. Pharmac.* **103**, 1196-1202 (1991).

17. Goldsmith, B. A. & Abrams, T. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9021-9025 (1991).
18. Eliot, L. S., Kandel, E. R., Siegelbaum, S. A. & Blumenfeld, H. *Nature* **361**, 634-637 (1993).
19. Mintz, I. & Korn, H. *J. Neurosci.* **11**, 3359-3370 (1991).
20. Carlson, J. H., Bergstrom, D. A., Weick, B. G. & Walters, J. R. *Synapse* **1**, 411-416 (1987).
21. Floran, B., Aceves, J., Sierra, A. & Martinez-Fong, D. *Neurosci. Lett.* **116**, 136-140 (1990).
22. Reubi, J.-C., Iversen, L. L. & Jessell, T. M. *Nature* **266**, 652-654 (1977).
23. Civelli, O., Bunzow, J. R. & Grandy, D. K. *Rev. Pharmac. Toxicol.* **33**, 281-307 (1993).
24. Weick, B. G. & Walters, J. R. *Brain Res.* **405**, 234-246 (1987).
25. Braun, A. R. & Chase, T. N. *Eur. J. Pharmac.* **131**, 301-306 (1986).
26. Robertson, G. S. & Robertson, H. A. *Trends pharmacol. Sci.* **8**, 295-299 (1987).
27. Lacey, M. G., Mercuri, N. B. & North, R. A. *J. Physiol.* **392**, 397-416 (1987).
28. Lacey, M. G., Mercuri, N. B. & North, R. A. *J. Physiol.* **401**, 437-453 (1988).
29. Williams, J. T., North, R. A., Shefner, S. A., Nishi, S. & Egan, T. M. *Neuroscience* **13**, 137-156 (1984).

ACKNOWLEDGEMENTS. We thank D. Bobker, J. Clements, N. Marrian and P. Osborne for their comments on the work and manuscript and Ciba-Geigy Basel for the gift of the CGP35348. This work was supported by grants from NIDA to J.T.W. and a C.J. Martin Fellowship from the Australian NH and MRC to D.L.C.

Synaptic vesicle fusion complex contains *unc-18* homologue bound to syntaxin

Yutaka Hata^{*†}, Clive A. Slaughter^{*} & Thomas C. Südhof^{*††}

^{*} Howard Hughes Medical Institute and [†] Department of Molecular Genetics, University of Texas Southwestern Medical School, Dallas, Texas 75235, USA

THREE synaptic proteins, syntaxin, SNAP-25 and synaptobrevin, were recently identified as targets of clostridial neurotoxins that irreversibly inhibit synaptic vesicle fusion¹⁻⁴. Experiments searching for membrane receptors for *N*-ethylmaleimide-sensitive fusion protein (NSF), which has an important role in membrane fusion, revealed an ATP-dependent interaction of the same three synaptic proteins with NSF and its soluble attachment proteins⁵. Thus, two independent approaches identify syntaxin, synaptobrevin and SNAP-25 as components of the synaptic vesicle fusion machinery, but their mode of action is unclear⁶. We have now discovered a brain protein of relative molecular mass 67,000 (67K) which binds stably to syntaxin. Amino-acid sequencing and complementary DNA cloning revealed that the 67K protein is encoded by the mammalian homologue of the *Caenorhabditis elegans* gene *unc-18*. In *C. elegans*, *unc-18* belongs to a group of genes defined by mutations with a paralytic phenotype and accumulations of acetylcholine, suggesting a defect in neurotransmitter release^{7,8}. The binding of the mammalian homologue of *unc-18* (Munc-18) to syntaxin requires the N terminus of syntaxin whereas that of SNAP-25 involves a more C-terminal sequence. Our data suggest that Munc-18 is a previously unidentified essential component of the synaptic vesicle fusion protein complex.

A recombinant protein containing the cytoplasmic domains of syntaxin A (also named HPC-1)^{9,10} fused to glutathione *S*-transferase (GST) was immobilized on glutathione-agarose beads to form an affinity column. Detergent-treated rat brain homogenate was applied to this column, washed and eluted with a sodium chloride step gradient, resulting in the purification of a single protein of *M_r* 67,000 (Fig. 1a). The 67K protein (named Munc-18, see below) binds tightly to GST-syntaxin and elutes from it in an essentially pure form. Two control experiments were performed to ensure that the 67K protein bound to the affinity column by a specific interaction with syntaxin. First, affinity purifications were carried out on glutathione-agarose under a variety of control conditions. Glutathione-agarose beads containing no attached proteins, glutathione-agarose beads containing GST alone, a GST-neurexin fusion protein (GST-Neu I) or GST-syntaxin were incubated with total brain homogenate and washed extensively and their total protein content was analysed by SDS-polyacrylamide gel electrophoresis (PAGE). Endogenous rat brain GSTs were purified on all of the beads, but the 67K protein bound only to GST-syntaxin (Fig. 1b). Second, a recombinant protein in which the syntaxin A sequence was fused to a six-histidine tag instead of GST was used for a similar affinity purification. Again, the 67K protein was the major protein purified (data not shown). Together, these data demonstrate that the 67K protein binds specifically and tightly to syntaxin A and constitutes the single major protein in brain which interacts with this component of the synaptic vesicle fusion complex.

Purified 67K protein was subjected to amino-acid sequencing for identification. An N-terminal sequence and internal sequences from tryptic peptides were obtained and used to isolate cDNA clones from rat and bovine brain. The translated amino-acid sequences of the rat and bovine cDNAs predict synthesis of a 67.6K hydrophilic protein with no apparent transmembrane region (Fig. 2). The rat and bovine sequences are very similar, showing only a single amino-acid substitution over 594 residues (leucine to valine at position 291). In contrast to the conservation of the coding region, the untranslated regions of the rat and bovine cDNAs show little homology, suggesting

[†] To whom correspondence should be addressed.

selective evolutionary pressure on the coding sequence of the protein (data not shown).

Sequence comparisons reveal that the 67K protein is highly homologous to the *unc-18* of *C. elegans*^{7,8} and a related gene from *Drosophila*¹¹, and weakly homologous to the products of several yeast genes implicated in the secretory pathway, in particular *sec1* (refs 12, 13) (Fig. 2). The 67K protein exhibits an overall identity of 57% with the *unc-18* product, suggesting that it represents the mammalian *unc-18* homologue (referred to as Munc-18). Mutations in the *unc-18* gene in *C. elegans* are characterized by paralysis, accumulations of acetylcholine and resistance to the acetylcholinesterase inhibitor aldicarb. A similar phenotype is observed in *C. elegans* only with mutations in a small group of genes⁸ that include the genes encoding synaptotagmin, a major synaptic vesicle Ca²⁺-binding protein^{14,15}, and the putative vesicular acetylcholine transporter¹⁶. The accumulations of acetylcholine in these mutants suggests that their phenotype is caused by a block in neurotransmitter release, indicating that the *unc-18* gene product is essential for neurotransmission. The previously described weak homology between the products of the *unc-18* and the yeast *sec1* genes¹⁷ can also be extended to the mammalian proteins that are slightly more homologous to *sec1* (23% overall identity) than is *unc-18* (21% identity) (Fig. 2). In addition, Munc-18 is weakly homologous to the yeast proteins slylp and slp1p, which are related to *sec1p* (ref. 13).

Thus, our results reveal that a protein encoded by the mammalian homologue of a *C. elegans* gene with an essential role in neurotransmission binds tightly to syntaxin, a component of the synaptic vesicle fusion machinery^{1,6}. To characterize this interaction further, we investigated its possible modulation.

Ca²⁺, ATP-γS, or GTP-γS have no measurable effect on the binding of Munc-18 to GST-syntaxin (Fig. 3a). However, treatment of the brain homogenate with *N*-ethylmaleimide (NEM), a sulphhydryl-reactive reagent which interrupts membrane fusion events, inhibits binding, as does prior freezing of the brain homogenate. This result suggests that the binding of Munc-18 to syntaxin is not regulated in the same way as the binding of the complex between NSF and α/β/γ-SNAPs to syntaxin/SNAP-25/synaptobrevin⁵, but that it is sensitive to perturbations in protein structure produced by NEM or freezing.

To investigate the stoichiometry of the Munc-18-syntaxin complex, glycerol gradient centrifugations were used. Syntaxin was cleaved from GST by thrombin and purified. Glycerol gradient centrifugations of purified syntaxin, purified Munc-18 and the Munc-18-syntaxin complex were performed. As judged by the values of their sedimentation constants, Munc-18 and the cytoplasmic domain of syntaxin migrate as monomers whereas the Munc-18-syntaxin complex has a sedimentation constant value corresponding to a 1:1 complex (Fig. 3b). In the experiment shown, the sample with the Munc-18-syntaxin complex contains an excess of syntaxin, resulting in the presence of monomeric syntaxin (Fig. 3b, closed arrows) and Munc-18-syntaxin complexes (open arrows). These results suggest that Munc-18 forms a stable complex with syntaxin, probably in a 1:1 ratio.

Syntaxin, SNAP-25 and synaptobrevin were identified as targets for clostridial neurotoxins¹⁻⁴ and as membrane receptors for the NSF-SNAP complex⁵, suggesting that these proteins are part of the synaptic vesicle fusion machinery. Syntaxin, SNAP-25 and synaptobrevin interact directly with each other and/or with the NSF-SNAP complex and possibly also with

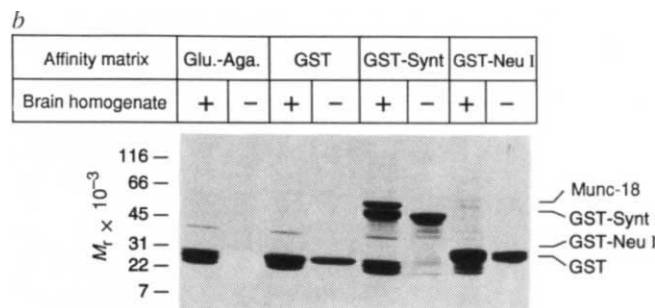
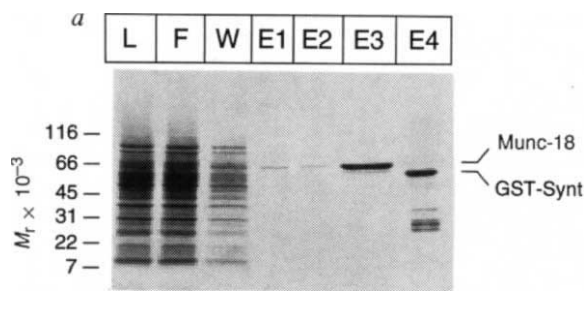


FIG. 1 Identification of a 67K brain protein that binds to syntaxin. *a*, Purification of a 67K brain protein (Munc-18) by affinity chromatography on immobilized syntaxin. Total rat brain homogenate was subjected to affinity chromatography on a GST-syntaxin fusion protein (GST-Synt) containing the cytoplasmic sequence of syntaxin A (HPC-1)^{9,10} (residues 1–261) bound to glutathione-agarose. Fractions were analysed by SDS-PAGE and Coomassie blue staining: L, column loading material; F, flow-through; W, Ca²⁺ wash; E1, Mg²⁺/EGTA eluent; E2, EDTA eluent in 0.1 M NaCl; E3, EDTA eluent in 1 M NaCl; E4, eluent with SDS-PAGE sample buffer. Migration positions of the 67K protein (Munc-18, see Fig. 2) and of GST-Synt are indicated on the right. The low *M_r* proteins eluted in sample buffer (E4) represent endogenous rat brain GSTs that bind to glutathione-agarose. *b*, Specificity of the binding of the 67K protein (Munc-18) to syntaxin. The picture shows a Coomassie blue-stained SDS-polyacrylamide gel of incubations of glutathione-agarose containing: no attached recombinant protein (labelled 'Glu.-Aga.'), recombinant GST, GST-Synt or GST-neurexin I fusion protein ('GST-Neu I'). Incubations were carried out in parallel with rat brain homogenate (left lanes) or control buffer (right lanes); positions of recombinant GST-fusion proteins are indicated on the right.

METHODS. Using standard procedures^{24,25}, the cytoplasmic domain of syntaxin A was cloned by the polymerase chain reaction from rat brain cDNA (oligonucleotide sequences: CGCGAATCCCGCGAGCATGAAG-GACCGAAC and GCGAAGCTTATGCCTTCTCTGGTACTTGAC), transferred into the bacterial expression vector pGEX-KG (ref. 26), sequenced, expressed and purified on glutathione-agarose²⁷. Other GST-protein used (Figs 3 and 4) were obtained similarly²⁸. To prepare total solubilized brain homogenate, frozen rat brains were homogenized in 4 mM

HEPES/NaOH pH 7.4, 0.1 g l⁻¹ phenylmethylsulphonyl fluoride (PMSF), extracted for 4 h at 4 °C after addition of an equal volume of 4 mM HEPES/NaOH pH 7.4, 0.1 g l⁻¹ PMSF, 0.2 M NaCl, 2% NP-40, 2 mM EDTA and centrifuged (100,000g for 30 min at 4 °C; yield: 1 g protein from 10 rats in 50 ml). For affinity chromatography (*a*), rat brain homogenate was precleared by incubation with 2 ml glutathione-agarose (6 h at 4 °C) and centrifugation (800g for 2 min). After addition of 3.5 mM CaCl₂ and MgCl₂, the homogenate was applied to a 1-ml glutathione-agarose column containing 6 mg GST-Synt and pre-equilibrated with buffer A (50 mM Tris/HCl pH 8.0, 0.1 M NaCl, 2.5 mM MgCl₂) containing 0.5% NP-40 and 2.5 mM CaCl₂. The column was washed with 20 ml buffer A containing 0.5% CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulphonate) and 2.5 mM CaCl₂ ('W' in *a*) and sequentially eluted with 10 ml of: buffer A containing 0.5% CHAPS and 5 mM EGTA (E1); 50 mM Tris/HCl pH 8.0, 0.1 M NaCl, 0.5% CHAPS and 5.0 mM EDTA (E2); 1 M NaCl, 0.5% CHAPS and 5 mM EDTA (E3); and 1.5 ml SDS-PAGE sample buffer (E4) (yield: 0.8 mg Munc-18 per g protein). For the batch affinity chromatography procedure (*b*), 50 μl glutathione-agarose without attached proteins or containing 0.5 nmol of the indicated GST-proteins were incubated overnight at 4 °C with 2 ml total brain homogenate (40 mg protein) in buffer B (4 mM HEPES/NaOH pH 7.4, 0.1 M NaCl, 1% NP-40, 1 mM EDTA, 3.5 mM CaCl₂, 3.5 mM MgCl₂) or in buffer B only. The glutathione-agarose was recovered by centrifugation (800g for 2 min), washed six times with 1.2 ml buffer A, resuspended in 0.18 ml SDS-PAGE sample buffer, and 40 μl was analysed by SDS-PAGE. Numbers on the left of the figures indicate positions of *M_r* markers.

FIG. 2 Structures of rat and bovine 67K proteins (Munc-18) determined by cDNA cloning: homologies to *C. elegans unc-18* and yeast *sec1* gene products. The amino-acid sequences of the rat and bovine 67K proteins as deduced from cDNA sequences (top two lines) are aligned with the sequences of the *C. elegans unc-18* gene product⁸ and of yeast *sec1*¹². Sequences are shown in single-letter amino-acid code and numbered on the right. Residues identical to the rat sequence are shaded, revealing an overall identity of 57% of the rat 67K protein with the *unc-18* gene product and identifying it as mammalian *unc-18* homologue (Munc-18). *Sec1*p is 23% identical to rat and bovine Munc-18 and 21% identical to the *unc-18* product¹⁷. Locations of amino-acid sequences determined from the purified protein are indicated by lines above the sequence (N-T, N-terminal sequence; #5, 6, 9, 17A, 17B, 17C and 18, sequences of tryptic peptides).

METHODS. Amino-acid sequencing of the purified 67K protein (Fig. 1) was performed as described²⁹. Full-length cDNA clones were isolated from rat and bovine brain cDNA libraries using oligonucleotides predicted by the amino-acid sequences and fully sequenced^{24,25}. In addition to the alignment shown here, a weak homology of Munc-18 to S1y1p and S1p1p, similar to the homologies of these proteins to *Sec1*p¹³, was detected (not shown). The nucleotide sequences of the cDNA clones reported here are deposited in Genbank. The accession number of rat *munc-18* is L26087 and that of bovine *munc-18* is L26088.

C. e.	MQNVKHSYLACSSSSCCGFSLFLSLGHAHQITLSPSFAYTLLIFALLVSDKSEIALSRHRTHTYIFLPISSVHPSL	79
	N-T	#17A
RAT	MAPIGKAVVGEKIMHDVKKVKKKGE--WKVLVDQSLMRMLSSCCCKMTDITTEGIVTIDINK--REEPLSPLEAVY	75
BOV	MAPIGKAVVGEKIMHDVKKVKKKGE--WKVLVDQSLMRMLSSCCCKMTDITTEGIVTIDINK--REEPLSPLEAVY	75
C. e.	ELLFPVFMHMKELNDVIRPLKKGDSAMNVLVDTLAMRLSSCGKMHNIHEEGITIVTIDINK--REEPLSPLEAVY	157
SEC1	MSDLIELQRYNLGLVNLQIETKNNKFLIIIDKTVETILSYLFLTLQELLNNVTSVDLSDPTKRGQSSIEATY	73
	#17B	#18
RAT	LITPSEKSVHSLISDFKDPPTAKYRAHVFFPDSGPDALFN--ELVKSRAAKVKTITTEINIAFLPYEQVYS--LDSADS	152
BOV	LITPSEKSVHSLISDFKDPPTAKYRAHVFFPDSGPDALFN--ELVKSRAAKVKTITTEINIAFLPYEQVYS--LDSADS	152
C. e.	LIAFTAESIDKLIQDTYCARNL--YKCAHVFFTEAGSDQLFS--TLKSAARFIKTKIEMIAFTPYEQVFN--LDSPT	232
SEC1	ILEPTKYNINCIDADPHVRFP--KYRCHIRFLPLGLTNPIQFFQSKRYIAQNLESFKFELGFFVKESQFFETLQMEHS	151
	#17C	#18
RAT	FQSFYSPHKAQKMPILERLAQIATLCATKREYPAVRYGEYK-----DNALLAQITQDKLDAYKA	214
BOV	FQSFYSPHKAQKMPILERLAQIATLCATKREYPAVRYGEYK-----DNALLAQITQDKLDAYKA	214
C. e.	FFLYYNAQKQGLTSLNLERIAQIATVATLGEYSRLRYADFE-----RNVELCHLVEQKLDAYKA	294
SEC1	LQVFNNNNKALIPITNVRKIVGSLVSLGVITGEYPIVRYVSNPVEEDARNGNAVNANSILRSIANAFQIAIDTYAR	230
	#17C	#18
RAT	DDPTH--GEGFDKARSQLLILDRGDFPSSPVLHETLQAMSYDLLPIENDVYK--YETSGIGEARVKE-----VLLDEDD	286
BOV	DDPTH--GEGFDKARSQLLILDRGDFPSSPVLHETLQAMSYDLLPIENDVYK--YETSGIGEARVKE-----VLLDEDD	286
C. e.	DDPSM--GEGADKARSQLLILDRGDAITFLHETLQAMCYDLLGIENDVYK--YETSGSDENLEKE-----VLLDEND	366
SEC1	NNDFPPQNTERRPSILIIIDRTLDFFAIIHDFSYQAMAYD--LVANVTQKDIHYSAENAEAGEQEKVSKLVLDLYP	308
	#17C	#18
RAT	LWIAVRHKHIAEVSQEVTRSLKDFSSSK--RMNTGKTTMRDLSQMLKMPQYQKLSKYSTHLEAEDCKHMQGIVD	363
BOV	LWIAVRHKHIAEVSQEVTRSLKDFSSSK--RMNTGKTTMRDLSQMLKMPQYQKLSKYSTHLEAEDCKHMQGIVD	363
C. e.	LWIAVRHKHIAEVSQEVTRSLKDFSSSK--RMNTGKTTMRDLSQMLKMPQYQKLSKYSTHLEAEDCKHMQGIVD	444
SEC1	DWIDELKQHIMDAVEYIQGRIKELIAKNFLVDRSNVKNMTDLISVVAHLKDFDEERKRIILHKTIVDEGLGENAERKL	387
	#17C	#18
RAT	-KLCRVEQDLAM--GTDAGEKIKDPKRAIVPILLDANVSTYDKIRIILLYIFLNGKITEENLNKIQHAQIPPEDESEI	440
BOV	-KLCRVEQDLAM--GTDAGEKIKDPKRAIVPILLDANVSTYDKIRIILLYIFLNGKITEENLNKIQHAQIPPEDESEI	440
C. e.	-KLCRVEQDLST--GTDAGEKVRDAMKLVPIIDPAVRCEDRLILLYIFLNGKITEENLNKIQHAQIPPEDESEI	521
SEC1	ADISAEQLNLSGFMDFSGEKKIKHIDDLIPALAMKEPTILDKLRYIIAYALFRGGIIEIDFKILNFIQVTHHEHNFQ	466
	#5	#9
RAT	TNMAHLGVPIVDSTLRRRSKPERKERISEQTYQLSWPTFIKIDMED--TIEDKLDTHYFYISTSSASFS-----	511
BOV	TNMAHLGVPIVDSTLRRRSKPERKERISEQTYQLSWPTFIKIDMED--TIEDKLDTHYFYISTSSASFS-----	511
C. e.	TNAAVGLNIVDTGRKKITPTTKERPEHQVYQSSKRVFVKIDIED--ATDERLDTHYFPLAGQVQVNGV-----	591
SEC1	QYKIFRNYDLIDFKLIK--DKPKDKFPQKFWHDTLVNDPNTYHSRFPVAVGNLSKVIANPILLSQEYFPYKDKPI	544
	#5	#9
RAT	-----TTAVSARYGIHWKNNK--PGYRSQFPLII-----FILGQVSLNEMKCAEYVTVQANGKWEVLIG	568
BOV	-----TTAVSARYGIHWKNNK--PGYRSQFPLII-----FILGQVSLNEMKCAEYVTVQANGKWEVLIG	568
C. e.	-----RAPASARYGQWKKERGQQSNYRSQFPLII-----YITGVTFSEMRACYEYVTAARKFWEVVG	649
SEC1	ELLNEEFQAGLANTSANSSSLRNPRHKAAMTKSSNIKNIPRQRFYFVIGGISIPEIKAAAYDQSNL--KNRDFIG	622
	#5	#9
RAT	STHILTPQKLLDTLKKLNKTDDEISS	594
BOV	STHILTPQKLLDTLKKLNKTDDEISS	594
C. e.	SDRIITDKKFLTNLRDLNKPRDI	672
SEC1	SDRIITDKKFLTNLRDLNKPRDI	724

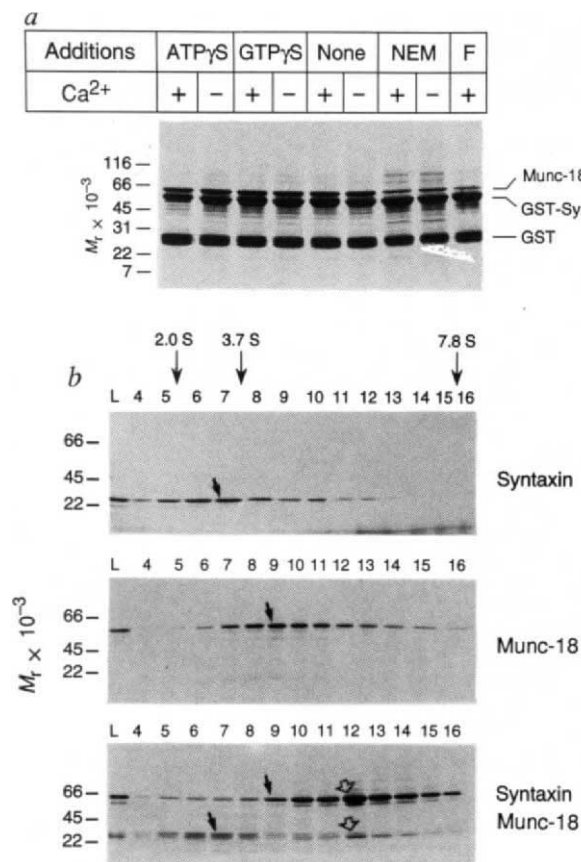


FIG. 3 Nature of the Munc-18-syntaxin complex. **a**, Dependence of Munc-18 binding to syntaxin on Ca²⁺, ATP-γS, GTP-γS, NEM and prior freezing of the brain homogenate. GST-syntaxin attached to glutathione-agarose was incubated with total rat brain homogenate in the presence or absence of Ca²⁺ (3.5 mM CaCl₂ or 5 mM EGTA) with 50 μM ATP-γS ('ATP-γS'), 50 μM GTP-γS ('GTP-γS'), or no further additions ('None'), or after pretreatment of the brain homogenate with 2 mM NEM for 2 h at 4 °C followed by the addition of 20 mM dithiothreitol ('NEM'), or after freezing the extracts at -20 °C for 24 h before use ('F'). Proteins bound to glutathione-agarose after extensive washing were analysed by SDS-PAGE and Coomassie blue staining. Positions of Munc-18, GST-Synt and endogenous brain GSTs purified on the glutathione-agarose are indicated on the right. **b**, Size analysis of the cytoplasmic domain of syntaxin, Munc-18 and the Munc-18-syntaxin complex by glycerol density gradient centrifugation. Fractions were analysed by SDS-PAGE and silver staining and are identified by numbers above each panel. Protein sizes are estimated³⁰ by comparison with known protein standards analysed on parallel gradients (2.0S, equine myoglobin (17.5K); 3.7S, chicken ovalbumin (44K); 7.8S, bovine γ-globulin (158K)). Closed arrows point to peak positions of uncomplexed syntaxin (2.8S, 31K) and Munc-18 (4.5S, 62K). Open arrows point to peak position of the Munc-18-syntaxin complex (5.9S, 94K).

METHODS. Experiments testing the effects of different treatments on Munc-18 binding to syntaxin were carried out as for Fig. 1b with additional buffer components or manipulations as stated. For the glycerol gradient centrifugations, GST-syntaxin either alone or complexed with Munc-18 (see Fig. 1) were cleaved with thrombin and purified²⁷. Partially purified Munc-18 was obtained as described in Fig. 1 legend. Samples (0.3 ml) adjusted to a glycerol concentration of 1% (w/v) were layered onto 11 ml of a 2–20% (w/v) linear glycerol gradient containing 4 mM HEPES/NaOH pH 7.4, 0.1 M NaCl, 0.5% NP-40, 2.5 mM CaCl₂ and 2.5 mM MgCl₂, and overlaid with 0.5 ml of 4 mM HEPES/NaOH pH 7.4. After centrifugation (15 h at 21,000g at 4 °C), 30 fractions were collected from the gradients and analysed by SDS-PAGE and silver staining²⁸.

synaptotagmin^{18,19}. Munc-18 binding to syntaxin may have been missed previously because freezing inhibits this interaction (Fig. 3a). Stoichiometric binding of SNAP-25, synaptotagmin and synaptobrevin to syntaxin was not observed on the Coomassie blue-stained gels in our experiments (Figs 1 and 3). However, such binding could be detected when more sensitive immuno-

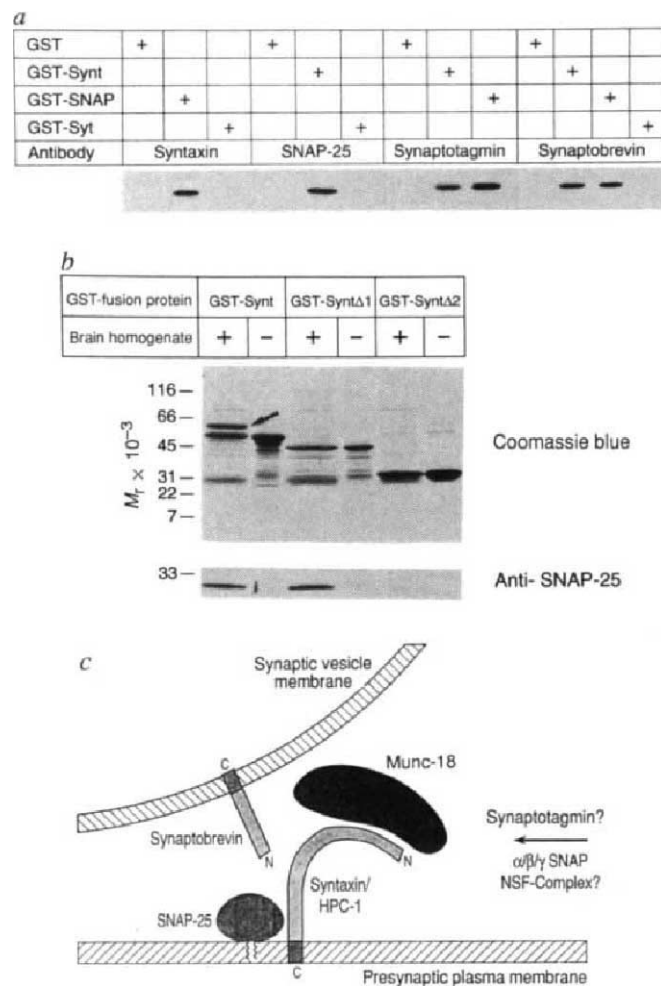


FIG. 4 Relation of the Munc-18-syntaxin complex to syntaxin binding of SNAP-25, synaptotagmin and synaptobrevin. **a**, Complex formation of syntaxin, SNAP-25, synaptotagmin and synaptobrevin. Total rat brain homogenates were incubated with the indicated GST-fusion proteins (GST alone, GST-syntaxin, GST-SNAP containing full-length SNAP-25, and GST-Syt, containing cytoplasmic domains of rat synaptotagmin I starting at residue 140; ref. 15). Proteins bound after extensive washing were analysed by SDS-PAGE and immunoblotting using the indicated antibodies. **b**, Sequence requirements of syntaxin for binding Munc-18 and SNAP-25. GST-fusion proteins containing the cytoplasmic domain of syntaxin (GST-Synt; residues 1–261), or deletions of residues 7–76 (GST-SyntΔ1), or of the C-terminal part of the syntaxin cytoplasmic domain (GST-SyntΔ2; residues 1–115) were attached to glutathione-agarose and incubated with rat brain homogenate or control buffer. Proteins bound to the agarose were analysed by SDS-PAGE followed by Coomassie blue staining (top) or immunoblotting with an antibody to SNAP-25 (bottom). Arrow points to position of Munc-18. **c**, Model of the interactions of proteins of the synaptic vesicle fusion complex. The interaction of syntaxin with SNAP-25, synaptobrevin and Munc-18 shown could be either simultaneous or sequential. Syntaxin also interacts with the NSF-SNAP complex⁵ and with synaptotagmin¹⁰, but the timing of these interactions and their stoichiometry are unclear. **METHODS**. Affinity purifications of proteins from brain homogenates using GST-fusion proteins were performed as for Fig. 1. Samples were analysed by SDS-PAGE followed by immunoblotting with the indicated antibodies described previously^{28,29}, and by Coomassie blue staining. Numbers on the left of the gels indicate positions of M_r markers.

blotting techniques were applied (Fig. 4a). For these experiments, recombinant fusion proteins of syntaxin, SNAP-25 and synaptobrevin with GST were used to bind proteins from total rat brain homogenates, demonstrating reciprocal but substoichiometric interactions between syntaxin, SNAP-25, synaptotagmin and synaptobrevin. These interactions were not modulated by Ca^{2+} , ATP-γS or GTP-γS, suggesting that they represent direct constitutive binding reactions (data not shown).

To determine whether interactions of syntaxin with different proteins are mediated by distinct sequences, two deletion mutants of GST-syntaxin were constructed and used for affinity purification of Munc-18 and SNAP-25. Deletion of residues 7–76 of the cytoplasmic domains of syntaxin completely abolishes Munc-18 binding but has no effect of the binding of SNAP-25 (Fig. 4b). Conversely, deletion of residues 116–261 abolishes binding of both Munc-18 and SNAP-25. These data suggest that Munc-18 binds to syntaxin at an N-terminal site that is distinct from the more C-terminal binding site for SNAP-25.

Our data demonstrate that syntaxin forms a tight stoichiometric complex with a 67K protein in brain which by amino-acid sequencing and cDNA cloning was identified as the mammalian homologue of the *C. elegans unc-18* gene, therefore referred to as Munc-18. Because syntaxin, together with SNAP-25 and synaptobrevin, is part of the synaptic vesicle fusion complex^{1–6}, Munc-18 probably represents a novel component of this complex. Figure 4c depicts a model of the protein-protein interactions operating in this complex in which Munc-18 is envisaged as playing a central role because of its stable interaction with syntaxin, although the exact roles of Munc-18 and other components of the complex in the fusion reaction are unknown. As *unc-18* seems to be essential for neurotransmission in *C. elegans*, Munc-18 is probably an essential component of the fusion complex. In addition to *unc-18*, Munc-18 is weakly homologous to yeast *sec1*, mutations in which inhibit exocytosis²⁰. *sec1* interacts with a family of genes named *SSo1* and *SSo2* whose products are related to syntaxin²¹ and, according to our data, are likely to interact physically with Sec1p. Multiple syntaxin isoforms are present in yeast and mammals which differ primarily in the N-terminal sequences and may act in different cellular fusion reactions^{22,23}. As Munc-18 binds to the N-terminus of syntaxin, it seems likely that mammalian cells will also contain a series of different Munc-18 isoforms, each of which may interact only with a specific isoform of syntaxin. Thus, Munc-18 and its isoforms could have a function in determining the specificity of intracellular fusion reactions. □

Received 20 October; accepted 5 November 1993.

- Link, E. et al. *Biochem. biophys. Res. Commun.* **189**, 1017–1023 (1992).
- Schiavo, G. et al. *Nature* **359**, 832–835 (1992).
- Blasi, J. et al. *Nature* **365**, 160–163 (1993).
- Blasi, J. et al. *EMBO J.* (in the press) **update?**
- Söllner, T. et al. *Nature* **362**, 318–324 (1993).
- Südhof, T. C., De Camilli, P., Niemann, H. & Jahn, R. *Cell* **75**, 1–4 (1993).
- Brenner, S. *Genetics* **77**, 71–94 (1974).
- Hosono, R. et al. *J. Neurochem.* **58**, 1517–1525 (1992).
- Inoue, I., Obata, K. & Akagawa, K. *J. Biol. Chem.* **267**, 10613–10619 (1992).
- Bennett, M. K., Calakos, N. & Scheller, R. H. *Science* **257**, 255–259 (1992).
- Salzberg, A., Cohen, N., Halachmi, N., Kimchie, A. & Leve, Z. *Development* **117**, 1309–1319 (1993).
- Aalto, M. K., Ruohonen, L., Hosono, K. & Keranen, S. *Yeast* **7**, 643–650 (1991).
- Aalto, M. K., Keranen, S. & Ronne, H. *Cell* **68**, 181–182 (1992).
- Nonet, M. L., Grundahl, K., Meyer, B. J. & Rand, J. B. *Cell* **73**, 1291–1305 (1993).
- Perin, M. S., Fried, V. A., Mignery, G. A., Jahn, R. & Südhof, T. C. *Nature* **345**, 260–261 (1990).
- Alfonso, A., Grundahl, K., Duerr, J. S., Han, H.-P. & Rand, J. B. *Science* **261**, 617–619 (1993).
- Pelham, H. R. B. *Nature* **364**, 582 (1993).
- Walch-Solimena, C. et al. *Neuron* (in the press).
- Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. *Cell* (in the press).
- Novick, P. & Scheckman, R. *Proc. natn. Acad. U.S.A.* **76**, 1858–1862 (1979).
- Aalto, M. K., Ronne, H. & Keranen, S. *EMBO J.* **12**, 4095–4104 (1993).
- Bennett, M. K. et al. *Cell* **74**, 863–873 (1993).
- Pelham, H. R. B. *Cell* **73**, 425–428 (1993).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1989).
- Südhof, T. C. et al. *Science* **245**, 1474–1480 (1989).

26. Guan, K. L. & Dixon, J. E. *Analyt. Biochem.* **192**, 262–267 (1991).
 27. Smith, D. B. & Johnson, K. S. *Gene* **67**, 31–40 (1988).
 28. Hata, Y., Davletov, B., Petrenko, A. G., Jahn, R. & Südhof, T. C. *Neuron* **10**, 307–315 (1993).
 29. Robinson, P. J. et al. *Nature* **365**, 163–166 (1993).
 30. Martin, R. G. & Ames, B. N. *J. biol. Chem.* **236**, 1372–1379 (1961).

ACKNOWLEDGEMENTS. We thank I. Leznicki, S. Afendis, C. Moomaw, E. Borowicz and A. Roth for technical assistance, and R. Jahn (New Haven, Connecticut) and M. Wilson (La Jolla, California) for antibodies. This work was supported by a postdoctoral fellowship to Y.H. from the Human Frontier Science Program.

GTP binding and hydrolysis by the signal recognition particle during initiation of protein translocation

Joshua D. Miller, Heike Wilhelm, Lila Gierasch*, Reid Gilmore† & Peter Walter

Department of Biochemistry and Biophysics,
University of California Medical School, San Francisco,
California 94143-0448, USA

* Department of Pharmacology,

University of Texas Southwestern Medical Center, Dallas,
Texas 75235-9041, USA

† Department of Biochemistry, University of Massachusetts,
Worcester, Massachusetts 01655-0001, USA

THE signal recognition particle (SRP) consists of one RNA and six protein subunits^{1,2}. The N-terminal domain of the 54K subunit contains a putative GTP-binding site, whereas the C-terminal domain binds signal sequences and SRP RNA^{3–7}. Binding of SRP to the signal sequence as it emerges from the ribosome creates a cytosolic targeting complex containing the nascent polypeptide chain, the translating ribosome, and SRP⁸. This complex is directed to the endoplasmic reticulum membrane as a result of its interaction with the SRP receptor^{9–11}, a membrane protein composed of two subunits, SR α and SR β , each of which also contains a GTP-binding domain^{12,13}. In the presence of GTP, SRP receptor binding to SRP causes the latter to dissociate from both the signal sequence and the ribosome^{13,14}. GTP is then hydrolysed so that SRP can be released from the SRP receptor and returned to the cytosol¹⁵. Here we show that the 54K subunit (*M*, 54,000) of SRP (SRP54) is a GTP-binding protein stabilized in a nucleotide-free

state by signal sequences, and that the SRP receptor both increases the affinity of SRP54 for GTP and activates its GTPase. We propose that nucleotide-mediated conformational changes in SRP54 regulate the release of signal sequences and the docking of ribosomes at the endoplasmic reticulum.

To analyse the role of GTP in protein targeting to the endoplasmic reticulum, (ER), we pursued the observation that the interaction of SRP with its receptor induces GTP hydrolysis¹⁶. Purified SRP had no detectable GTPase activity and the purified SRP receptor hydrolysed GTP only poorly, but SRP and its receptor together hydrolysed GTP about ten times faster than the receptor alone (Fig. 1).

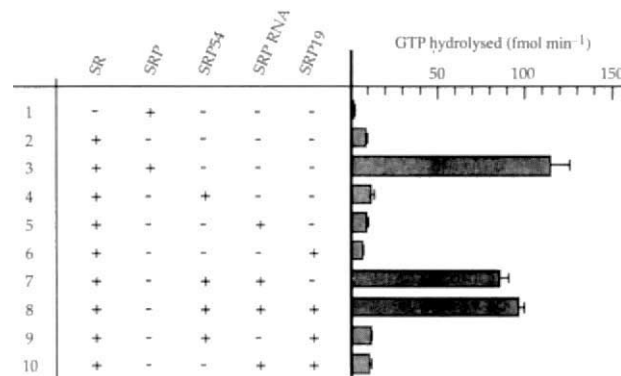
To determine which components of SRP interact with the SRP receptor to increase GTP hydrolysis, SRP was dissociated into its subunits under non-denaturing conditions. The dissociated proteins can be purified and reconstituted with SRP RNA to regenerate fully functional SRP¹⁷. Surprisingly, a partially reconstituted SRP containing only the 19K and 54K subunits and SRP RNA was almost as active as native SRP in the presence of SRP receptor (Fig. 1), but was inactive without it (not shown). Moreover, omission of SRP19, which stabilizes the binding of SRP54 to SRP RNA¹⁸, reduced activity only slightly (Fig. 1). SRP RNA and SRP54, however, were both essential (Fig. 1), indicating that the complex of SRP54 and SRP RNA is both necessary and sufficient to elicit GTP hydrolysis in conjunction with the SRP receptor. All subsequent analysis was carried out with this 'minimal' SRP [SRP(54/RNA)].

To determine whether SRP54, SR α or SR β catalyses the GTP hydrolysis, we monitored nucleotide binding to the proteins by ultraviolet crosslinking^{19,20}. This approach allowed us to detect the relatively low-affinity binding of these proteins to GTP and to measure GTP binding to each of the three GTP-binding proteins in the reaction independently. When SRP receptor was incubated with [α -³²P]GTP and crosslinked using ultraviolet radiation, both SR α and SR β were labelled (Fig. 2a, lane 1). Similarly, when SRP(54/RNA) was used, SRP54 was labelled (Fig. 2a, lane 2). SRP54 was crosslinked to the same extent when SRP RNA was omitted (not shown), and was the only SRP protein labelled when intact SRP was crosslinked, indicating that the labelling reaction was specific for GTP-binding proteins (not shown). SRP54, SR α and SR β must therefore be GTP-binding proteins, as predicted from their amino-acid sequences.

When SRP(54/RNA) and SRP receptor were mixed to stimulate GTP hydrolysis (Fig. 1, reaction 7), GTP crosslinking to SRP54 was dramatically stimulated (Fig. 2a, lane 3), but there was no significant change in crosslinking to either SR α or SR β .

FIG. 1 Stimulated GTPase activity of SRP and partially reconstituted SRPs. GTP hydrolysis rates are the average of three independent experiments; the standard deviation of the measurements is indicated. The reaction was linear with time over the period analysed. tRNA could not replace SRP RNA in this reaction. In the presence of the SRP receptor (SR), all partially reconstituted SRPs that contained both the SRP RNA and SRP54 were about equally active; that is, the additional presence of SRP68/72 and/or SRP9/14 had no effect on the reaction. In the absence of SR, purified SRP proteins, SRP RNA and all partially reconstituted SRPs were inactive.

METHODS. SRP and SRP receptor were purified as described^{28,29} as were the individual SRP components¹⁸. Partially reconstituted SRPs were formed by mixing components at a concentration of 500 nM each in 300 mM potassium acetate, 5 mM Mg(OOCH₃)₂, 25 mM HEPES pH 7.5, 0.01% Nikkol detergent, (octaethyleneglycol mono-*n*-dodecyl ether; Nikko Chemical, Tokyo), 1 mM dithiothreitol (DTT). After mixing, reactions were incubated for 10 min on ice, 10 min at 37 °C and then kept on ice until the GTPase reaction. GTPase reactions (20 μ l) contained 20 nM SR and/or either 20 nM SRP or 20 nM partially reconstituted SRPs in GTP hydrolysis buffer containing 50 mM KOOCH₃, 50 mM triethanolamine, pH 7.5, 2.5 mM Mg(OOCH₃)₂, 0.5% Nikkol detergent, 1 mM DTT. GTP 1 μ M included 0.5 mCi ml⁻¹ [γ -³²P]GTP (ICN). Reactions were incubated at 25 °C for 20 min and assayed by charcoal adsorption followed by Cerenkov counting.



There was no stimulation of GTP crosslinking to SRP54 when SRP RNA was omitted (not shown). Thus, GTP crosslinking to SRP54 (Fig. 2) and GTP hydrolysis (Fig. 1) are both stimulated by a functional interaction between SRP(54/RNA) and SRP receptor.

Most guanine nucleotide-binding proteins bind GDP tightly and are stimulated to bind GTP by specific guanine-nucleotide-releasing proteins (GNRPs), which decrease their affinity for GDP, creating a transiently empty nucleotide-binding site for GTP to enter²¹. To test whether SRP receptor was acting as a GNP, we tested its effect on the ability of SRP54 to bind GDP. SRP(54/RNA) was crosslinked to [α -³²P]GDP in the presence of different concentrations of unlabelled competitor nucleotide. The concentration of nucleotide required to inhibit the crosslinking of SRP54 is a measure of its apparent affinity for the protein. As expected, crosslinking of SRP54 to [α -³²P]GDP could be competed with unlabelled GDP (Fig. 2c, open diamonds). Surprisingly, addition of receptor to these reactions did not alter the concentration of GDP required to inhibit the crosslinking (Fig. 2c, filled diamonds), indicating that SRP does not function as a GNP to decrease the apparent affinity of SRP54 for GDP.

When unlabelled GTP was added to compete with [α -³²P]GDP for crosslinking, (Fig. 2d, triangles), roughly tenfold higher concentration of unlabelled nucleotide was required for half-

maximal inhibition (IC_{50}). In the absence of SRP receptor, the apparent affinity of SRP54 for GTP is thus ~10-fold lower than that for GDP. As SRP receptor stimulates GTP hydrolysis, a similar experiment with the receptor present would be difficult to interpret. We therefore used a non-hydrolysable GTP analogue, GMP-PNP, which has an apparent affinity for SRP54 which is threefold lower than GTP in the absence of the receptor (Fig. 2d, open squares). When SRP receptor was added to these reactions, the IC_{50} for GMP-PNP decreased ~90-fold (Fig. 2d, filled squares), indicating that the receptor substantially increased the apparent affinity of SRP54 for GTP. SRP receptor, therefore, does not act as a GNP to release GDP, but instead appears to be a 'guanine-nucleotide-loading protein' which promotes GTP binding by increasing the affinity of SRP54 for GTP.

As SRP receptor stimulates the binding of GMP-PNP to SRP54 under conditions that lead to GTP hydrolysis, SRP54 may be responsible for nucleotide hydrolysis. If so, the labelled nucleotide crosslinked to SRP54 (Fig. 2a, lane 3) may be hydrolysed to GDP. The experiment shown in Fig. 2a was therefore repeated with [γ -³²P]GTP. The ³²P-labelled γ -phosphate would be released upon hydrolysis, resulting in the loss of the radiolabel from the crosslinked protein. SR α , SR β and SRP54 were labelled with [γ -³²P]GTP to the same extent as they were with [α -³²P]GTP when reactions were done with either SRP receptor

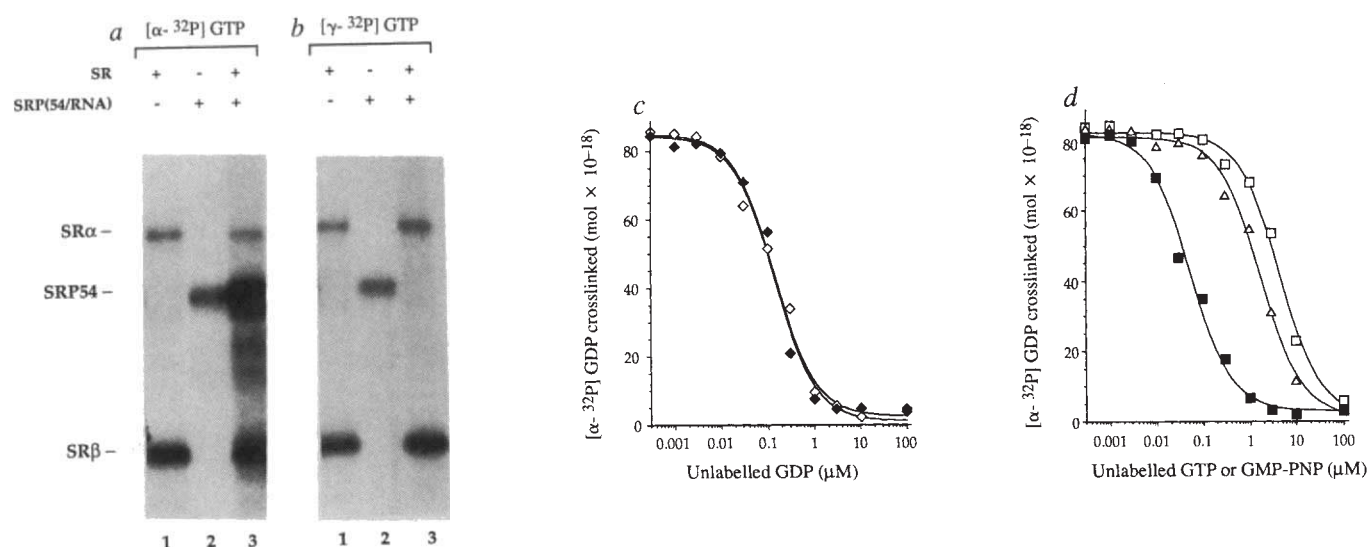


FIG. 2 SRP receptor both increases the affinity of SRP54 for GTP and functions as a GTPase-activating protein. *a*, SR (lane 1), SRP(54/RNA) (lane 2) or both complexes combined (lane 3) were ultraviolet-crosslinked to [α -³²P]GTP. *b*, Reactions were identical to those in *a*, except that [γ -³²P]labelled GTP was used. *c*, Crosslinking of radiolabelled GDP is competed with increasing amounts of unlabelled GDP in either the presence (◆) or absence (◇) of SR. The apparent K_i for GDP in the presence or absence of SR was 0.14 μ M. *d*, Crosslinking of radiolabelled GDP is competed for by increasing amounts of unlabelled GTP in the absence of SR (△) or GMP-PNP in the presence (■) or absence (□) of SR. The apparent K_i for GTP in the absence of SR was 1.7 μ M. This is consistent with the apparent K_d of SRP54 for GTP (2 μ M) as determined by crosslinking directly to labelled GTP. The apparent K_i s for GMP-PNP in the absence and in the presence of SR were 4.5 and 0.05 μ M, respectively.

METHODS. *a* and *b*, Reactions (20 μ l) containing 20 nM SRP(54/RNA) and/or SR were incubated for 20 min in GTP hydrolysis buffer at 25 °C. The GTP concentration was 0.3 μ M, including 0.5 mCi ml⁻¹ [α -³²P]GTP (Amersham) (*a* and *b*) or 0.5 mCi ml⁻¹ [γ -³²P]GTP (*c*). After 20 min, reactions were pipetted into plastic weigh boats and irradiated with ultraviolet light at 6,000 W per cm² (using eight Model G15T8 15W Germicidal lamps (General Electric) 6 cm from the sample) for 5 min on ice to crosslink covalently the bound radiolabelled nucleotide to the

protein¹⁹. Reactions were then precipitated with trichloroacetic acid to remove uncrosslinked label and analysed by SDS-PAGE followed by autoradiography. Ultraviolet crosslinking of labelled GTP to SRP54, SR α and SR β was saturable and specific for GTP, as it could be inhibited by an excess of unlabelled GTP but not by ATP or CTP. *c* and *d*, Reactions (20 μ l) contained 20 nM SRP(54/RNA) and/or SR in GTP hydrolysis buffer. All reactions also contained 0.1 μ M GDP including 0.5 mCi ml⁻¹ [α -³²P]GDP. Individual reactions were supplemented with unlabelled nucleotides to the concentrations indicated. Reactions were incubated for 4 h at 25 °C to reach equilibrium and then UV-crosslinked. Under saturating conditions, the level of crosslinking of [α -³²P]GDP to SRP54 was quantitatively identical to that of [α -³²P]GTP (not shown), indicating that crosslinking efficiencies are invariant for different nucleotides. At saturation, 4×10^{-4} mol nucleotide crosslinked to one mol SRP54. Crosslinked product was quantified using a PhosphorImager (Molecular Dynamics). Data points are experimental and the line is generated as a best fit to the equation: $B = B_{max}[1 - [I]/([I] + K_i(1 + ([S]/K_d)))]$ (a modification of equation III-5 described in ref. 30) using the program Kaleidagraph (Abelbeck Software) on a Macintosh II computer; B , is the amount of [α -³²P]GDP crosslinked to SRP54; B_{max} , amount of [α -³²P]GDP crosslinked to SRP54 in the absence of competitor; $[I]$, concentration of competitor; K_i , dissociation constant of competitor; K_d , dissociation constant of [α -³²P]GDP; $[S]$, concentration of [α -³²P]GDP.

or SRP(54/RNA) alone (compare Fig. 2a and b, lanes 1 and 2), confirming that neither component alone hydrolyses GTP significantly. In the reaction containing both receptor and SRP(54/RNA), however, the labelling of SRP54 was almost completely abolished (Fig. 2b, lane 3), whereas labelling of SR α and SR β was unchanged, indicating that the nucleotide crosslinked to SRP54 was GDP. SRP54 must thus hydrolyse GTP upon interaction with SRP receptor, and the receptor therefore functions not only as a guanine-nucleotide-loading protein, but also as a GTPase-activating protein.

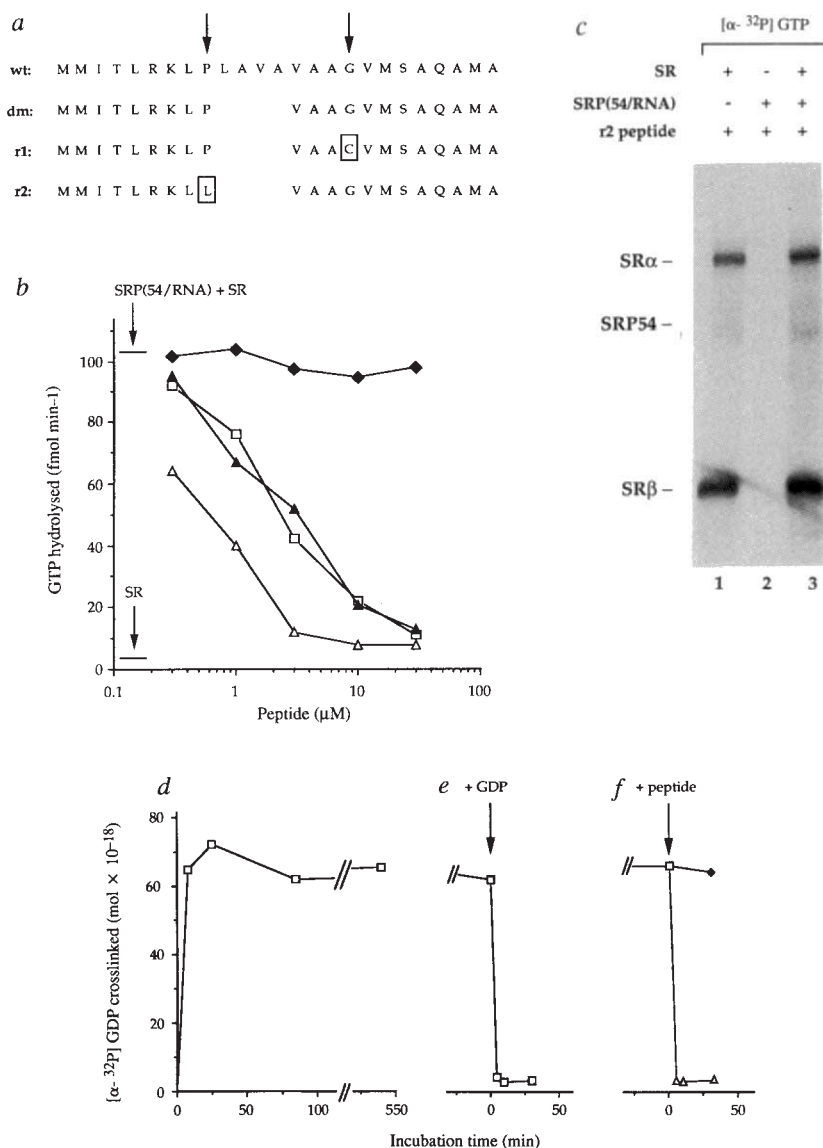
As SRP54 is bound to a signal sequence during the targeting reaction, we investigated whether a signal sequence could influence the receptor-dependent GTP hydrolysis reaction. Four synthetic peptides, derived from the signal sequence of the bacterial outer membrane protein LamB were tested (Fig. 3a): the wild-type signal sequence (wt), a deletion mutant (dm), which renders the signal sequence inactive *in vivo*, and two different second-site, single-amino-acid reversions that restore activity *in vivo* (r1 and r2)^{22,23}. The three peptides corresponding to functional signal sequences all strongly inhibited receptor-dependent GTP hydrolysis by SRP54, with IC₅₀s of 2, 2 and 0.4 μ M, respectively

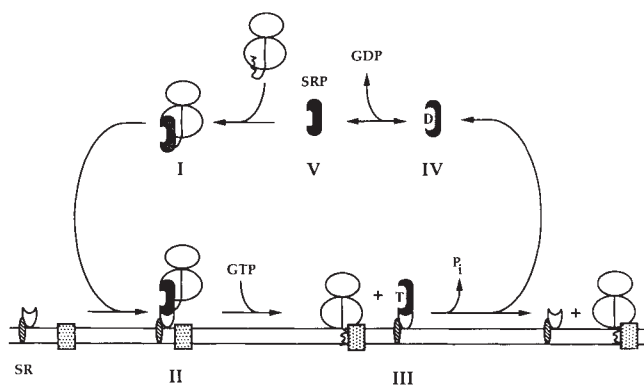
(Fig. 3b). The deletion mutant control peptide, however, did not affect the reaction even at 30 μ M, the highest concentration tested (Fig. 3b).

To determine whether the peptides were inhibiting GTP hydrolysis by blocking GTP binding to SRP54 or by blocking the hydrolysis step, an ultraviolet crosslinking experiment was done in the presence of the peptides. Crosslinking of [α -³²P]GTP to SRP54 was drastically reduced in the presence of a functional signal peptide, with or without SRP receptor (compare Figs 2a and 3c). GTP crosslinking to SR α and SR β was unaffected by the peptide (Fig. 3c), and the deletion mutant peptide did not inhibit GTP crosslinking to SRP54, SR α or SR β (not shown). Binding of a functional signal peptide to SRP54 must thus prevent GTP binding and so inhibit GTP hydrolysis.

The inhibition of GTP binding to SRP54 mediated by the signal peptides might reflect either stabilization of a nucleotide-free state or stabilization of a bound nucleotide (co-purified with SRP54), preventing labelled GTP from entering the occupied binding site. To discriminate between these mechanisms, [α -³²P]GDP was prebound to SRP54 as shown in Fig. 3d, and then an excess of unlabelled GDP was added (Fig. 3e).

FIG. 3 Functional signal peptides inhibit GTP binding to and hydrolysis by SRP54. a, Synthetic peptides used in the GTPase assay. The wild-type (wt) peptide corresponds to the signal sequence of the *E. coli* LamB protein. The LamB signal sequence functions efficiently in a mammalian *in vitro* translocation system³¹, and the synthetic peptides used here inhibit *in vitro* protein translocation across the membrane of *E. coli* inverted vesicles³². The synthetic wild-type peptide can readily adopt an α -helical conformation when analysed by circular dichroism spectroscopy²³. The deletion mutant peptide does not form an α -helix²³, and the peptide does not function as a signal peptide *in vivo*²². In the second site revertants r1 and r2, either the glycine or the proline residue are changed to a different amino acid. Both the r1 and r2 peptides regain the ability to form an α -helix and function as signal sequences *in vivo*^{22,23}. b, GTP hydrolysis reactions (as in Fig. 1) containing SRP(54/RNA) were supplemented with increasing concentrations of the synthetic wt ($\square$), dm ($\blacklozenge$), r1 ($\blacktriangle$) or r2 ($\triangle$) peptides. The degree of GTP hydrolysis in the absence of signal peptide [SRP(54/RNA) + SR] and the basal level of hydrolysis by SR are indicated (SR). The peptides were added to the reactions from a 150 μ M stock solution in water, the concentration of which was determined by amino-acid analysis. c, Reactions were identical to those for Fig. 2a, except that the r2 signal peptide was added to 4 μ M. d, [α -³²P]GDP was mixed with SRP(54/RNA) in GTP hydrolysis buffer. Aliquots were removed for UV crosslinking at the times indicated. The amount of GDP crosslinked to SRP54 was determined after SDS-PAGE by quantitation with a PhosphorImager. e, SRP(54/RNA) was incubated with GDP as in d. At time zero, unlabelled GDP is added to a final concentration of 10 μ M. Aliquots were removed for UV crosslinking at the indicated times to monitor the dissociation of the prebound [α -³²P]GDP. f, SRP(54/RNA) was incubated with GDP as in d. At time zero, the r2 ($\triangle$) or dm ($\blacklozenge$) peptide were added to final concentrations of 4 μ M. All reactions contained 20 nM SRP(54/RNA) in GTP hydrolysis buffer and 0.1 μ M GDP including 0.5 mCi⁻¹ ml [α -³²P]GDP. UV crosslinking and analysis were done as for Fig. 2.





Crosslinking of labelled GDP to SRP54 was radically diminished even at the earliest time point after addition of unlabelled GDP, indicating that the radiolabelled GDP dissociated rapidly. Results were identical when the r2 signal peptide was added to SRP54 containing prebound [α - 32 P]GDP (Fig. 3f, triangles), indicating that the signal peptide does not stabilize a GDP-bound state. Addition of the control deletion mutant peptide (Fig. 3f, diamond) did not reduce GDP crosslinking. A functional signal sequence thus appears to stabilize the nucleotide-free form of SRP54, although it remains a formal possibility that a signal sequence bound to SRP54 still allows nucleotide binding but causes a conformational change in the protein that completely disrupts ultraviolet crosslinking to the bound nucleotide.

Taken together, our data indicate that the guanine-nucleotide-bound state of SRP54 can be influenced by at least two ligands, the signal peptide and the SRP receptor, suggesting that the GTPase domain of SRP54 integrates information from the nascent polypeptide and the ER. We propose a model in which the occupancy of the SRP54 guanine-nucleotide-binding site defines discrete steps in a cycle of GTP binding and hydrolysis which operates during protein targeting to the ER (Fig. 4).

In the first step, signal sequence binding to SRP54 initiates formation of the targeting complex (Fig. 4, steps V→I). Previous studies have shown that guanine nucleotide and signal sequence bind to structurally separate domains on SRP54 (refs 5, 6). Moreover, chemical modification of the GTP-binding domain prevents signal sequence binding²⁴. Our data suggest that binding of a signal sequence to one domain stabilizes an empty nucleotide binding site in the other. Thus, the two domains can communicate bidirectionally and the binding of a signal peptide and guanine nucleotide (GTP or GDP) to SRP54 may be mutually exclusive. We therefore propose that the targeting complex arrives at the ER membrane with SRP54 held in a nucleotide-free state by the signal sequence (Fig. 4, step I→II), when an SRP receptor-catalysed conformational change in SRP54 (Fig. 4, II) stimulates GTP binding and concomitantly reduces its

FIG. 4 Model depicting the proposed role of GTP binding to and hydrolysis by SRP54 during the initiation of protein translocation. As discussed in the text, SRP54 is proposed to cycle between three forms with respect to bound nucleotide: a nucleotide-free or 'empty' state, a GTP-liganded (T), and GDP-liganded (D) state. The translocation apparatus is indicated by the stippled box.

affinity for the bound signal sequence (Fig. 4, III). As purified receptor is not sufficient to stimulate GTP binding to SRP54 in the presence of a signal peptide (Fig. 3c), this step may require some part of the targeting complex not included in the reconstituted assays, or a constituent of the membrane. This additional factor may be necessary to ensure that the signal sequence is not released from SRP unless essential translocation site components are available.

After GTP binding to SRP54 and release of the signal sequence, SRP and its receptor dissociate from the translocation apparatus as a stable complex^{14,15} (Fig. 4, III). SRP receptor-stimulated hydrolysis of GTP by SRP54 then ensues, allowing SRP to dissociate from its receptor and return to the cytosol (Fig. 4, IV). Given the rapid dissociation of GDP from SRP54 (Figs 2c and 3e), nucleotide-bound SRP54 probably exists in equilibrium with empty SRP54 (Fig. 4, step IV→V), which can bind a signal sequence to initiate another round of targeting (Fig. 4, step V→I).

According to this model, SRP54 closely resembles other GTPases (such as the trimeric G proteins or EF-Tu), in that interconversion between the different nucleotide-bound states causes the GTPase to interact successively with its effectors^{25,26}. Whereas the crucial conformational switch for other well characterized GTPases is the interconversion between the GTP- and GDP-bound states, interconversion between the empty state and the GTP-bound state seems to be the important switch for SRP54.

This model attributes no function to the other GTPase domains in SR α and SR β . It is not yet known whether the GTPase domain of SR β is required for receptor function, but the GTP-bound form of SR α is necessary for the targeting reaction to progress through the cycle shown in Fig. 4 (ref. 27), suggesting that SR α may undergo GTP binding and hydrolysis like SRP54. It is possible that, just as SRP recruits ribosomes carrying nascent chains from the cytosol to the translocation site, so the SRP receptor may recruit essential components of the translocation apparatus from the plane of the membrane. □

Received 17 May; accepted 14 September 1993.

- Walter, P. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7112–7116 (1980).
- Walter, P. & Blobel, G. *Nature* **99**, 691–698 (1982).
- Bernstein, H. D. et al. *Nature* **340**, 482–486 (1989).
- Römisch, K. et al. *Nature* **340**, 478–482 (1989).
- Zopf, D., Bernstein, H. D., Johnson, A. E. & Walter, P. *EMBO J.* **9**, 4511–4517 (1990).
- Römisch, K., Webb, J., Lingelbach, K., Gausepohl, H. & Dobberstein, B. *J. Cell Biol.* **111**, 1793–1802 (1990).
- High, S. & Dobberstein, B. *J. Cell Biol.* **113**, 229–233 (1991).
- Nunnari, J. & Walter, P. *Curr. Opin. Cell Biol.* **4**, 573–580 (1992).
- Gilmore, R., Blobel, G. & Walter, P. *J. Cell Biol.* **95**, 463–469 (1992).
- Gilmore, R., Walter, P. & Blobel, G. *J. Cell Biol.* **95**, 470–477 (1992).
- Meyer, D. I., Krause, E. & Dobberstein, B. *Nature* **297**, 647–650 (1982).
- Tajima, S., Lauffer, L., Rath, V. L. & Walter, P. *J. Cell Biol.* **103**, 1167–1178 (1986).
- Connolly, T. & Gilmore, R. *Cell* **57**, 599–610 (1989).
- Gilmore, R. & Blobel, G. *Cell* **35**, 677–685 (1983).
- Connolly, T., Rapiejko, P. J. & Gilmore, R. *Science* **252**, 1171–1173 (1991).
- Connolly, T. & Gilmore, R. *J. Cell. Biol.* (in the press).
- Walter, P. & Blobel, G. *Cell* **34**, 525–533 (1983).
- Siegel, V. & Walter, P. *J. Cell Biol.* **100**, 1913–1921 (1985).

- Nath, J. P., Eagle, G. R. & Himes, R. H. *Biochemistry* **24**, 1555–1560 (1985).
- Pashev, I. G., Dimitrov, S. I. & Angelov, D. *Trends Biochem. Sci.* **16**, 323–326 (1991).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117–127 (1991).
- Emr, S. & Silhavy, T. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4599–4603 (1983).
- McKnight, C. J., Briggs, M. S. & Gierasch, L. M. *J. Biol. Chem.* **264**, 17293–17297 (1989).
- Lütcke, H., High, S., Römisch, K., Ashford, A. & Dobberstein, B. *EMBO J.* **11**, 1543–1551 (1992).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
- Gilman, A. G. *Rev. Biochem.* **56**, 615–649 (1987).
- Rapiejko, P. J. & Gilmore, R. *J. Cell Biol.* **117**, 493–503 (1992).
- Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 682–691 (1983).
- Migliaccio, G., Nicchitta, C. V. & Blobel, G. *J. Cell Biol.* **117**, 15–25 (1992).
- Segel, I. H. *Enzyme Kinetics: Behavior and Analysis of Steady State and Rapid Equilibrium Enzyme Systems* 105 (Wiley-Interscience, New York, 1975).
- Watanabe, M., Hunt, J. F. & Blobel, G. *Nature* **323**, 71–73 (1986).
- Chen, L., Tai, P. C., Briggs, M. S. & Gierasch, L. M. *J. Biol. Chem.* **262**, 1427–1429 (1987).

ACKNOWLEDGEMENTS. We thank H. Bourne and J. Nunnari for discussion, members of P.W.'s laboratory for critically reading the manuscript, and C. Nicchitta for technical assistance. This work was supported by grants from the NIH and the Human Frontiers Science Foundation (to P.W.).

Defective mitosis due to a mutation in the gene for a fission yeast 26S protease subunit

Colin Gordon, Gordon McGurk, Patrick Dillon^{*†},
Craig Rosen^{*†} & Nicholas D. Hastie

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

^{*} Department of Gene Regulation, Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110-1199, USA

[†] Human Genome Sciences, 9620 Medical Center Drive, Rockville, Maryland 20850, USA

WE have isolated a mutant, *mts2*, in the fission yeast *Schizosaccharomyces pombe* which is defective in chromosome segregation. The predicted amino-acid sequence of the cloned *mts2*⁺ gene product is 75% identical to the S4 subunit of the human 26S ATP/ubiquitin-dependent protease¹. The human S4 subunit complementary DNA expressed from an *S. pombe* expression plasmid can rescue an *S. pombe mts2* gene disruption. Both observations demonstrate that the *mts2*⁺ gene is the *S. pombe* homologue of the human S4 subunit. In addition, we provide genetic evidence for a physical interaction between the S4 and the related S7 subunit in the 26S multiprotein protease. We show that polyubiquitin-conjugated proteins accumulate in the *mts2* mutant at the restrictive temperature, demonstrating that the mutant has an *in vivo* defect in the ubiquitin-dependent proteolysis pathway. Finally, the phenotype for the *mts2* mutant indicates that protein degradation by the 26S protease is essential not for entry into but for the completion of mitosis.

We carried out a screen to isolate mutants defective in microtubule function in the fission yeast *S. pombe*. Screening was for mutants that were resistant to the mitotic poison methylbenzylcarbamylate (MBC), which also displayed a temperature-sensitive phenotype. Mutants in five genes were identified. Genetic analysis demonstrated that for each mutant a single gene mutation was responsible for both the temperature sensitivity and resistance to methylbenzylcarbamylate. In heterozygous diploids, all the mutants were recessive for the conditional lethal phenotype and semi-dominant for the drug-resistance phenotype. We call the mutants *mts* (for methylbenzylcarbamylate-resistant, temperature-sensitive).

One mutant, *mts2*, was characterized further. After four hours at the restrictive temperature, approximately 25% of cells were found to condense the DNA into three chromosome-like structures (Fig. 1b). Associated with the condensed DNA was a short metaphase spindle (Fig. 1a). This phenotype was transient, as on continued incubation at the restrictive temperature the DNA decondensed and the spindle disassembled. A septum formed in ~40% of the cells, dividing the cells into a nucleate and an anucleate cell (Fig. 1c). The phenotype is similar to some of the previously described *cut* mutants². Consistent with this phenotype, histone H1 kinase activity increased in extracts obtained from *mts2* cells incubated at the restrictive temperature (Fig. 1d). FACS analysis (Fig. 2) demonstrates that cells arrested with a G2 content of DNA. Thus the conclusion from the cytological and flow cytometry analysis is that the *mts2* mutant is defective in chromosome segregation.

The *mts2*⁺ gene was cloned by complementation of the conditional lethal phenotype and shown to be the authentic gene by homologous integration. One allele of the *mts2*⁺ gene was disrupted with the *S. pombe ura4*⁺ gene in a diploid, and polymerase chain reaction analysis was used to confirm that the correct gene had been disrupted (data not shown). When this was sporulated, tetrads segregated 2:2 viable:non-viable. All the viable segregants were uracil auxotrophs, consistent with the *mts2*⁺ gene

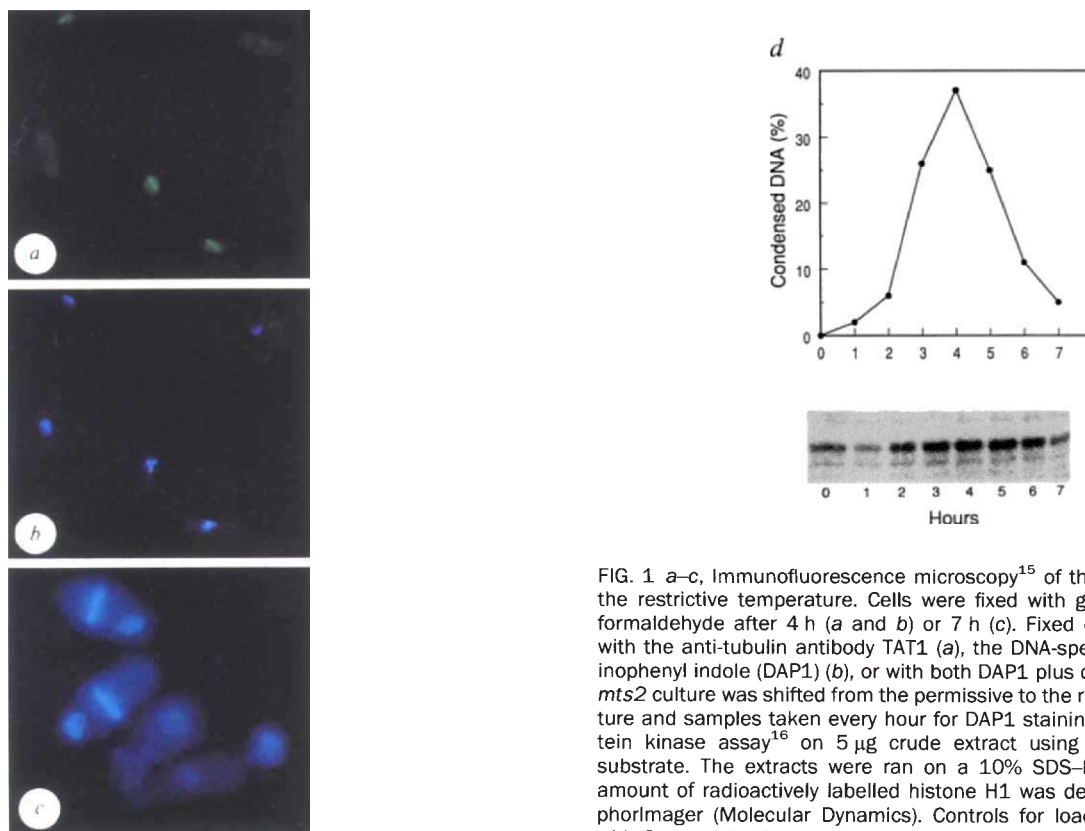


FIG. 1 a–c, Immunofluorescence microscopy¹⁵ of the *mts2* mutant at the restrictive temperature. Cells were fixed with glutaraldehyde and formaldehyde after 4 h (a) and 7 h (b) or 7 h (c). Fixed cells were stained with the anti-tubulin antibody TAT1 (a), the DNA-specific stain diaminophenyl indole (DAP1) (b), or with both DAP1 plus calcofluor (c). d, An *mts2* culture was shifted from the permissive to the restrictive temperature and samples taken every hour for DAP1 staining and p34^{cdc2} protein kinase assay¹⁶ on 5 µg crude extract using histone H1 as a substrate. The extracts were ran on a 10% SDS-PAGE gel and the amount of radioactively labelled histone H1 was detected on a PhosphorImager (Molecular Dynamics). Controls for loading were stained with Coomassie blue.

product being essential for growth. The *mts2⁺* gene was sequenced and found to encode a predicted protein of 448 amino acids. The predicted amino-acid sequence was 73% identical to the amino-acid sequence of the HeLa S4 subunit of the 26S protease (Fig. 3a). The HeLa S4 cDNA was subcloned into the *S. pombe* expression vector pREP1 (ref. 3). This construct was used to demonstrate that the human S4 cDNA could rescue both the *mts2* conditional mutant and the *mts2* null allele (Table 1). The structural similarity and the functional rescue of a *mts2* null by the HeLa cell S4 cDNA demonstrates that the *S. pombe mts2* gene is the fission yeast homologue of the human gene.

Before publication of the HeLa S4 cDNA sequence, we isolated a mouse cDNA (from a mouse cDNA library cloned in an *S. pombe* expression vector) that could rescue the conditional lethality of the *mts2* mutant. Surprisingly, sequencing this cDNA demonstrated that it was not the mouse homologue of the *mts2* gene but of a related gene, the human *MSS1* gene⁴ (Fig. 3b). Additional experiments showed that this cDNA could not rescue an *mts2* null allele (Table 1). This provides genetic evidence that there might be an interaction between the S4 (*mts2*) and the *MSS1* gene products in which the *MSS1* gene product might stabilize the temperature-sensitive *mts2* protein, allowing it to function at the restrictive temperature. This is consistent with biochemical evidence from human cells which shows that both proteins are present in the 26S protease complex¹. The accompanying letter⁵ demonstrates that the sequence of the *S. cerevisiae* gene *CIM5* is very similar to the human *MSS1* gene. Interestingly, a conditional mutation in this gene also resulted in a defect in chromosome segregation.

The 26S protease complex is thought to be involved in the ATP/ubiquitin-dependent proteolysis of target proteins⁶. To investigate whether the *mts2* mutant is defective in ubiquitin-dependent proteolysis *in vivo*, protein extracts were prepared from *mts2* and wild-type cells that had been incubated at the restrictive temperature. These extracts were probed with an antibody against ubiquitin protein conjugates⁷. As shown in Fig. 3d, the serum recognized a mixed population of high-molecular-weight proteins in extracts prepared from the *mts2* mutant but not from wild-type cells incubated at the restrictive temperature. In addition, extracts prepared from *nuc2* cells, which also arrest in metaphase but encode a nuclear scaffold protein², do not show an increase in the abundance of high-molecular-weight species when probed with the antibody (data not shown). This provides direct evidence that ubiquitin-dependent degradation is defective in the *mts2* mutant.

In *Xenopus* the regulatory protein cyclin B has to be degraded for transition from metaphase to anaphase to occur⁸ and the ubiquitin pathway is thought to be involved in cyclin B proteolysis⁹. A simple explanation for the phenotype observed in the *mts2* mutant is that cyclin B is no longer being degraded in its highly regulated manner. To investigate this possibility, protein extracts obtained from *mts2* cells incubated at the restrictive temperature were probed with an antibody against *S. pombe* cyclin B (Cdc 13). As shown in Fig. 3e, we found no obvious differences.

The result shown in Fig. 3e raises at least three possibilities for the *in vivo* targets of the 26S protease whose lack of degradation result in a defect in mitosis. First, Cdc 13 could still be the target, but only a specific intracellular subset, which although still stabilized in *mts2* mutant cells would not be visualized by probing total cell extract.

TABLE 1 Rescue of *mts2* mutants by the S4 and MSS1 cDNAs

	pS4	pMSS1	pmts2 ⁺	pREP1
<i>mts2-ts</i>	+	+	+	—
<i>mts2-null</i>	+	—	+	—

pREP1 is a *S. pombe* expression vector with no insert and is a negative control. pS4, pMSS1 and pmts2⁺ are the human S4, mouse MSS1 and *S. pombe mts2⁺* cDNAs cloned into pREP1. These plasmids were used to transform an *mts2 ts* mutant to determine whether they could rescue the conditional lethality when grown at the restrictive temperature, and an *mts2Δura4⁺/mts2⁺;leu1.32/leu1.32;ura4-D18/ura4-D18;ade6M216/ade6M210;h⁺/h* heterozygous diploid which was then sporulated to determine whether the plasmid could rescue an *mts2* null allele. A plus sign indicates growth; a minus sign indicates no growth.

Second, two other cyclin genes, *cig1* and *cig2*, have recently been identified in *S. pombe*^{10,11}; perhaps these or as-yet unidentified cyclin proteins could be the *in vivo* target. It is notable that in *S. cerevisiae* four G2 specific cyclins have been reported^{12,13}. A third possibility is that some other non-cyclin-like protein (for example, see ref. 14) needs to be degraded for mitosis to progress. Further experiments will be required to investigate each of these possibilities. □

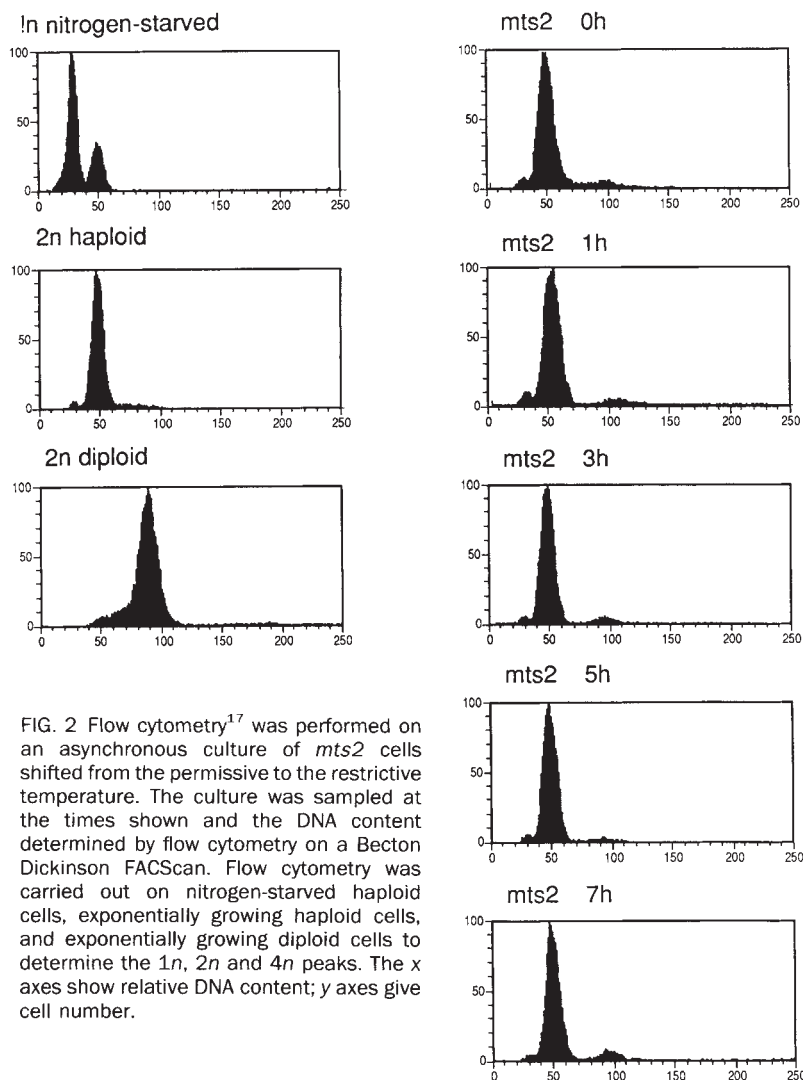


FIG. 2 Flow cytometry¹⁷ was performed on an asynchronous culture of *mts2* cells shifted from the permissive to the restrictive temperature. The culture was sampled at the times shown and the DNA content determined by flow cytometry on a Becton Dickinson FACScan. Flow cytometry was carried out on nitrogen-starved haploid cells, exponentially growing haploid cells, and exponentially growing diploid cells to determine the 1n, 2n and 4n peaks. The x axes show relative DNA content; y axes give cell number.

***S. cerevisiae* 26S protease mutants arrest cell division in G2/metaphase**

Michel Ghislain*, Andor Udvardy† & Carl Mann

Service de Biochimie et Génétique Moléculaire, Département de Biologie Cellulaire et Moléculaire, Centre d'Etudes de Saclay, F-91191 Gif-sur-Yvette Cedex, France

† Institute of Biochemistry-Biological Research Center, Hungarian Academy of Sciences, H-6701 Szeged, PO Box 521, Hungary

* Present address: Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

We isolated two mutants from the yeast *Saccharomyces cerevisiae*, *cim3-1* and *cim5-1*, that arrest cell division in G2/metaphase at 37 °C. CIM3 (identical to SUG1; ref. 1) and CIM5 are similar to each other and are members of a family of putative ATPases that have been proposed to be 26S protease subunits². We show here that CIM5 is the functional yeast homologue of the human MSS1 protein³ and that homologues of CIM3 and CIM5 are present in a highly purified preparation of the *Drosophila* 26S protease⁴. The short-lived ubiquitin–proline- β -galactosidase fusion protein is stabilized in *cim* mutants, but Leu- β -galactosidase is not. The CLB2 and CLB3 cyclins also accumulate in the *cim* mutants. Thus the 26S protease is required *in vivo* for the degradation of ubiquitinated substrates and for anaphase chromosome separation.

The CDC28 protein kinase of *S. cerevisiae* is required in G1 phase for entry into a new cycle and in G2 phase for mitosis⁵. We isolated *cim* (for co-lethal in mitosis) mutations that result in lethality when in combination with the G2/mitosis-specific *cdc28-1N* mutation^{6,7}, in the hope of identifying downstream targets of the kinase. We describe here the characterization of the temperature-sensitive *cim3-1 CDC28*⁺ and *cim5-1 CDC28*⁺ mutants. After incubation at 37 °C for 4 h, about 90% of the *cim3-1* cells and 70% of the *cim5-1* cells had a large bud containing a single nucleus (Fig. 1a). Most of the *cim3-1* and *cim5-1* cells arrested division with replicated DNA (Fig. 1b) and a short, intranuclear spindle characteristic of *S. cerevisiae* cells arrested in G2 or mitotic metaphase⁸ (Fig. 1c).

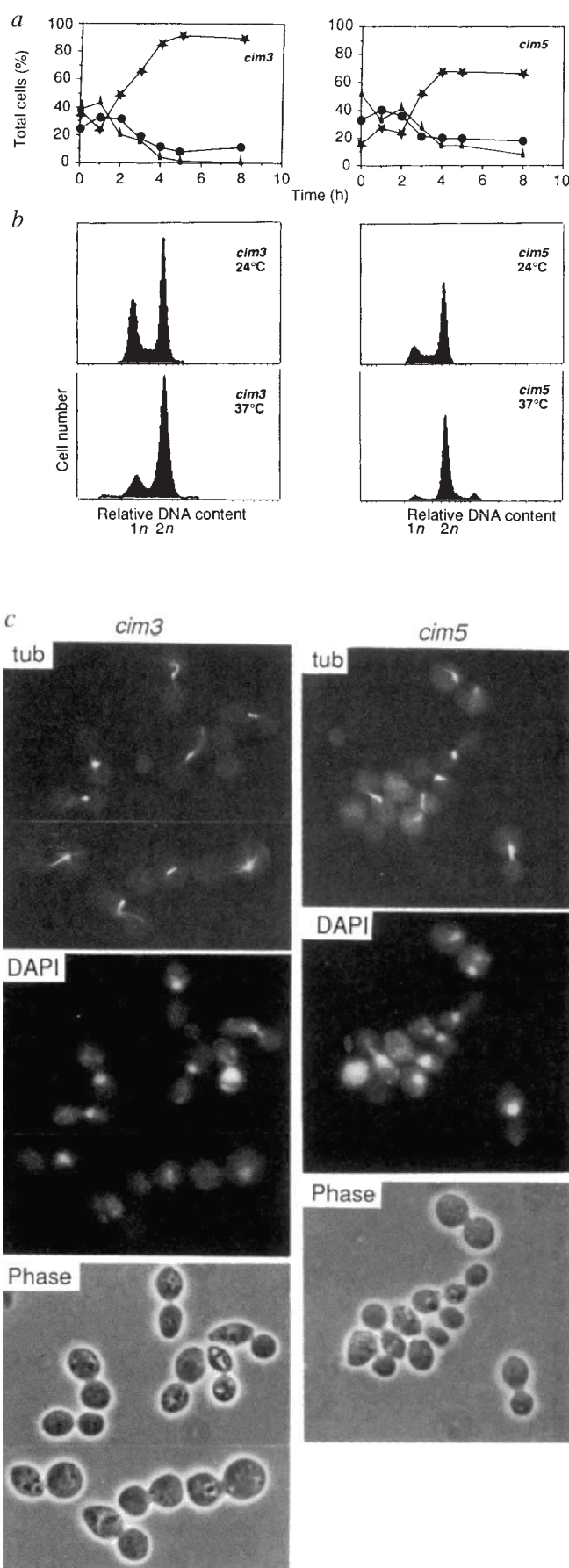


FIG. 1 *cim3-1* and *cim5-1* mutants arrest cell division in G2/metaphase. **a**, *cim3-1* and *cim5-1* cells arrest division with large buds. *cim* mutants growing exponentially at 24 °C were shifted to 37 °C. The percentage of total cells without buds (●), with a small bud (▲) and with a large bud (★) was determined as a function of time at 37 °C. The *cim3-1* mutant arrested division within one cell cycle, whereas the *cim5-1* mutant underwent two cell cycles during the 8-h incubation at 37 °C. **b**, *cim3-1* and *cim5-1* cells arrest division with a 2n DNA content as seen by flow cytometry²¹ of cells grown at 24 °C or at 37 °C for 8 h. The positions of the 1n (G1) and 2n (G2 or mitotic) DNA peaks are indicated. The *cim3-1* and *cim5-1* mutants have an abnormally high fraction of cells with a G2 DNA content even at 24 °C compared to the isogenic wild type (not shown). **c**, Microtubules (tub) were observed in *cim3-1* and *cim5-1* diploid cells after 6 h of incubation at 37 °C by immunofluorescence with YOL1/34 anti-tubulin antibody as described²². 87% of the *cim3-1* cells and 55% of the *cim5-1* cells were arrested with a short, intranuclear spindle ($n=140$). The *cim5-1* cells undergo some residual cell division at 37 °C that may account for the smaller number of cells containing an intranuclear spindle. Some cells show a single pole of fluorescence because all spindles were not in the same focal plane. An exponentially growing population of isogenic wild-type diploid cells at 37 °C was composed of 24% of cells with a short intranuclear spindle, 16% of cells with a long intranuclear spindle and 60% of cells with no visible intranuclear spindle. DNA was visualized in the same cells with diamidinophenyl indole (DAPI)²²; cells were observed in phase contrast.

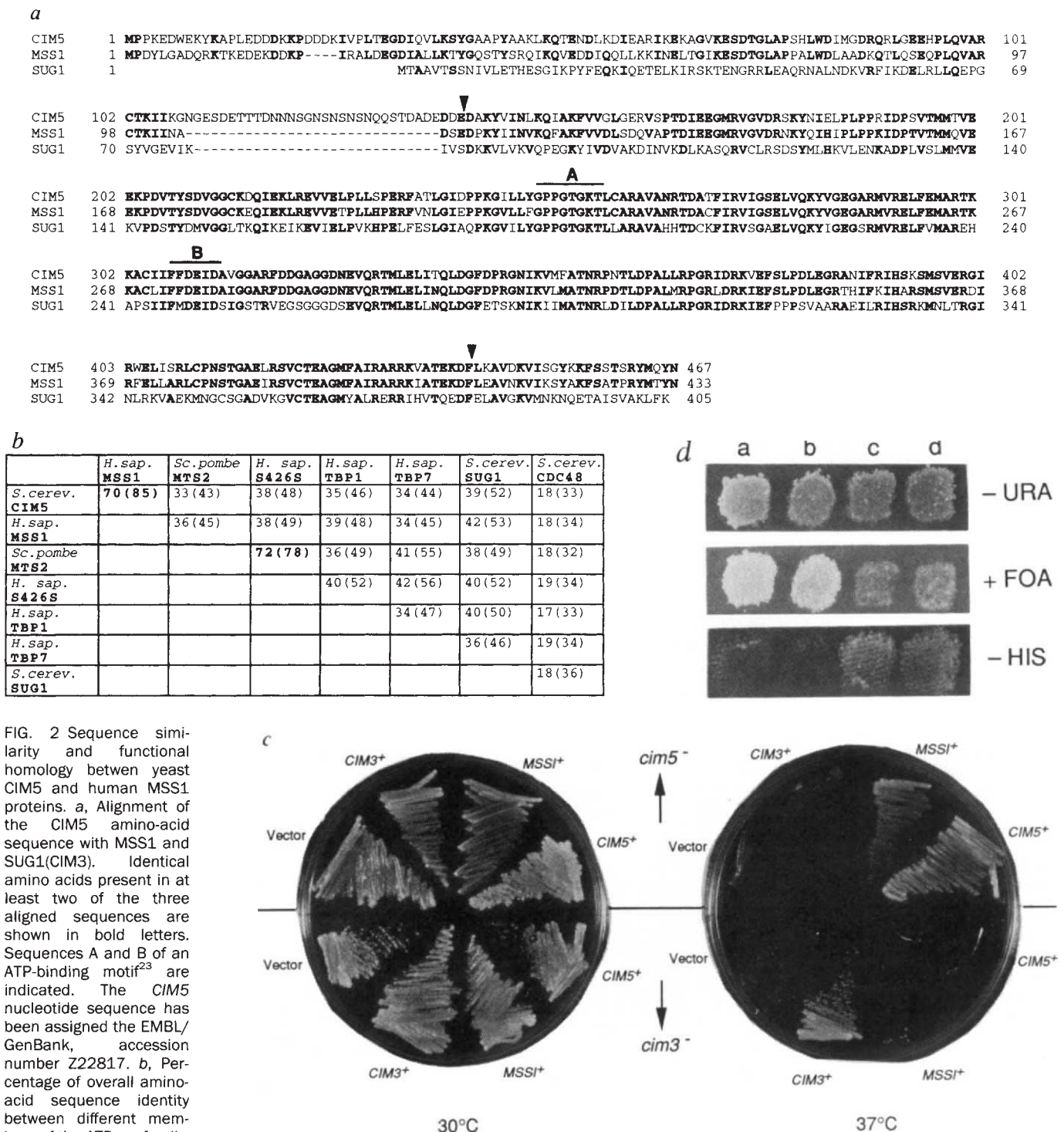


FIG. 2 Sequence similarity and functional homology between yeast CIM5 and human MSS1 proteins. **a**, Alignment of the CIM5 amino-acid sequence with MSS1 and SUG1(CIM3). Identical amino acids present in at least two of the three aligned sequences are shown in bold letters. Sequences A and B of an ATP-binding motif²³ are indicated. The CIM5 nucleotide sequence has been assigned the EMBL/GenBank accession number Z22817. **b**, Percentage of overall amino-acid sequence identity between different members of the ATPase family, including S426S (fourth subunit of the human 26S protease²), TBP1 (ref. 24) (a putative Tat-binding protein) and TBP7 (ref. 24). In parentheses is given the percentage identity confined to a highly conserved 300-amino-acid domain (arrowheads in **a**) surrounding the ATP-binding motif. CDC48 is one of a family of functionally diverse proteins containing two apparent ATP-binding domains²⁵. Two protein pairs (CIM5/MSS1 and MTS2 (ref. 19)/S426S) are highly similar and functionally homologous. **c**, MSS1 suppresses the temperature-sensitive defect of the *cim5-1* mutant. *cim3-1* and *cim5-1* were transformed with multicopy plasmids carrying either the MSS1, CIM3 or CIM5 genes or the vector alone, and incubated at 30 °C (left) and 37 °C (right). The human MSS1 cDNA expressed from plasmid pNV11 (ref. 3) complements the *cim5-1* but not the *cim3-1* temperature-sensitive defect. Multiple copies of the CIM3 gene cannot suppress the *cim5-1* mutation nor can multiple copies of CIM5 suppress the *cim3-1* mutation. **d**, The MSS1 cDNA complements the lethality of

a *cim5::HIS3* deletion mutant. A *cim5::HIS3/CIM5 ura3/ura3 his3/his3* pNV11 (*URA3*)-MSS1³ diploid strain was sporulated and a representative tetrad in which each spore had received pNV11(*URA3*)-MSS1 was patched onto YPD plates and then replica plated to synthetic medium lacking uracil (-URA), lacking histidine (-HIS) or containing 5-fluoro-orotic acid (+FOA)²⁶. The FOA inhibits the growth of *URA*⁺ cells containing pNV11(*URA3*)-MSS1. The CIM5^{his3} haploids (**a** and **b**) could lose the MSS1 plasmid and grow on the FOA plates, whereas the *cim5::HIS3* haploids (**c** and **d**) were unable to grow on the FOA plates, demonstrating their dependence on the MSS1 plasmid for viability. The *cim5::HIS3* mutation was made by replacing an 813-bp *Afl*II-*Bgl*III internal fragment of CIM5 with a 1.7-kb *HIS3* fragment. This construction deletes sequences coding for 271 of the 467 amino acids of CIM5 and creates a recessive lethal mutation.

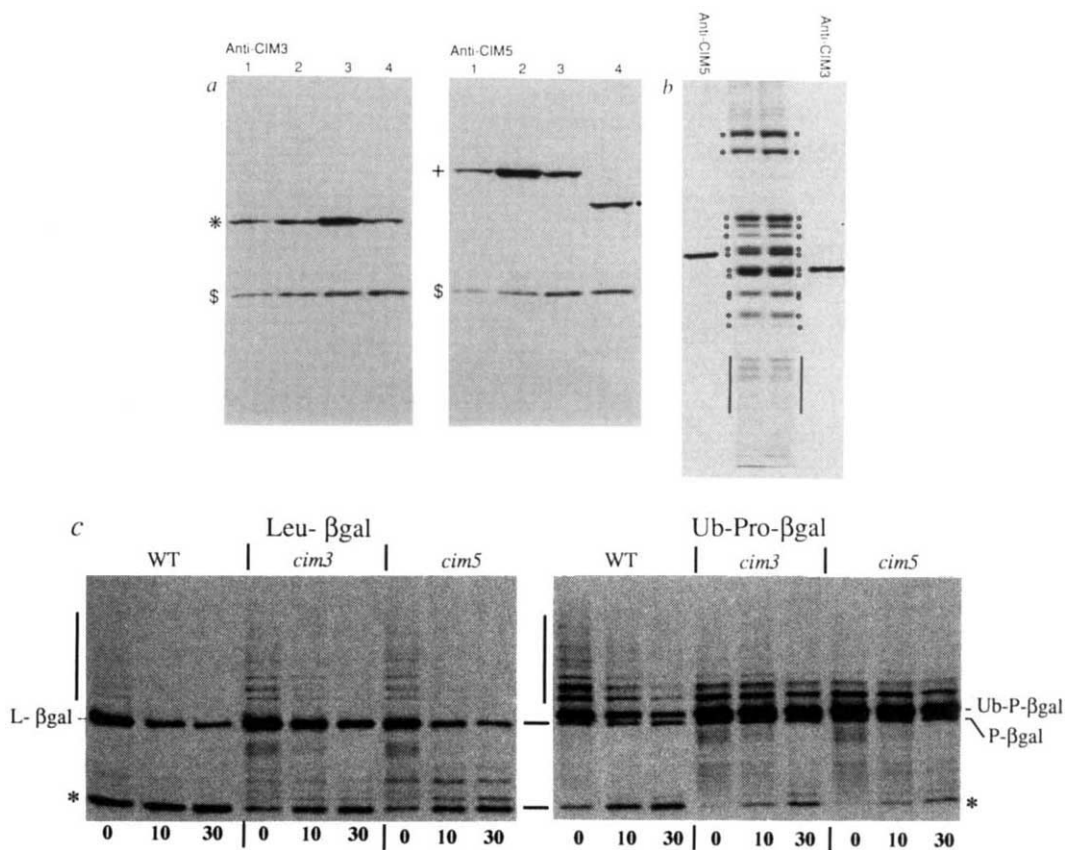
Wild-type *CIM3* and *CIM5* genes were cloned and sequenced. *CIM3* is identical to *SUG1* (suppressor of *gal4*)¹. *CIM5* is an essential gene (Fig. 2d) encoding a novel 467-amino-acid polypeptide. *CIM3*(*SUG1*) and *CIM5* are about 40% identical to each other and to a class of proteins characterized by a 300-amino-acid domain containing an ATP-binding motif (Fig. 2a and b). *CIM5* is 70% identical to the human *MSS1* protein³ and the *MSS1* cDNA complemented the temperature-sensitive phenotype of the *cim5-1* mutation as well as the lethality of a *cim5::HIS3* mutation (Fig. 2c and d).

Affinity-purified anti-*CIM3* and anti-*CIM5* antibodies specifically recognized their corresponding antigens in yeast whole-cell extracts (Fig. 3a). Anti-*CIM5* antibodies also cross-reacted with the human *MSS1* protein expressed instead of *CIM5*. We tested our anti-*CIM* antibodies for cross-reaction with a nearly homogeneous preparation of the *Drosophila* 26S protease⁴. Proteins of $M_r \sim 43K$ and $48K$ were recognized by the anti-*CIM3* and anti-*CIM5* antibodies, respectively (Fig. 3b). These proteins are part of a complex (the μ particle) that forms the 26S protease when combined with the 20S proteasome and a third, unincorporated factor⁴. This result supports the

contention² that these putative ATPases are 26S protease subunits.

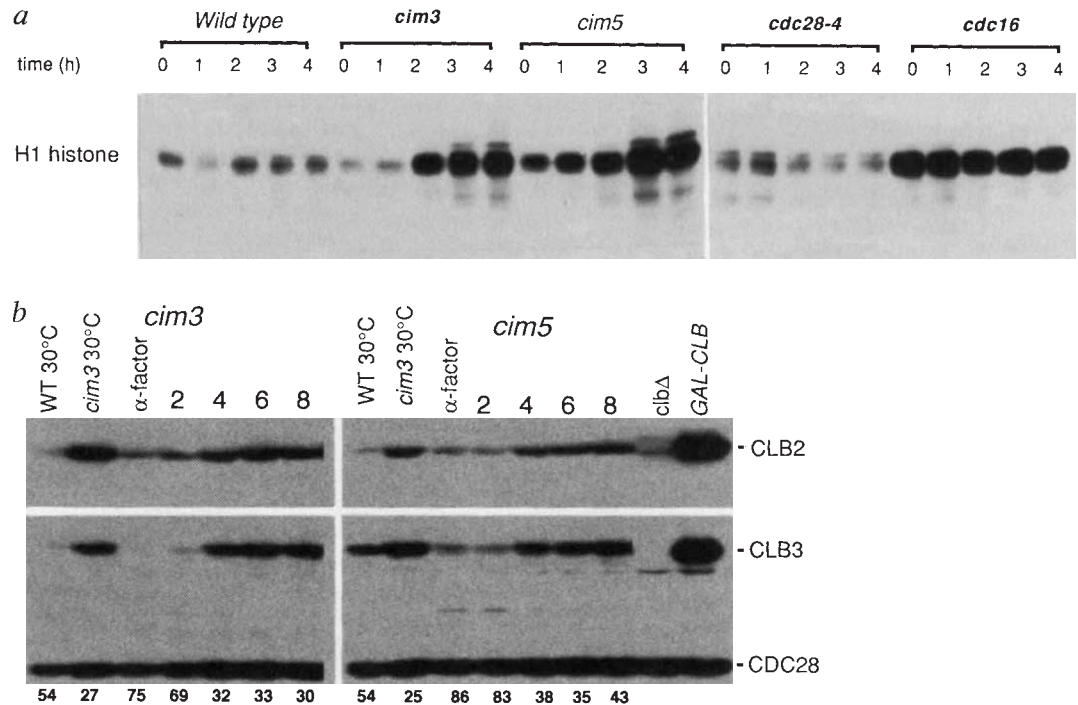
The 26S protease, but not its 20S proteasome core, degrades ubiquitin-conjugated proteins *in vitro* in an ATP-dependent reaction⁹. We therefore examined the stability of known substrates of the ubiquitin degradation pathway in *cim3-1* and *cim5-1*. Ubiquitin-Leu- β -galactosidase and ubiquitin-Arg- β -galactosidase are rapidly deubiquitinated *in vivo*¹⁰ and targeted to the N-end rule degradation pathway¹¹. Ubiquitin-Pro- β -galactosidase (Ub-Pro- β gal) is slowly deubiquitinated and is degraded through a different ubiquitin-dependent pathway¹². Pulse-chase analysis indicated that Ub-Pro- β gal was stabilized in the *cim* mutants relative to the isogenic wild-type strain (Fig. 3c). The half-life of Ub-Pro- β gal was estimated at 9, 21 and 45 min for wild type, *cim3-1* and *cim5-1* strains, respectively. However, degradation of Leu- β gal (Fig. 3c) and Arg- β gal (not shown) was not significantly different in the *cim* mutants compared with wild-type. Thus, *CIM3* and *CIM5* are required for the proteolysis of Ub-Pro- β gal, but may not be involved in the degradation of N-end rule substrates. In contrast, the *pre1-1* and *pre2-2* mutants, affecting essential subunits of the yeast 20S proteasome,

FIG. 3 *CIM3*(*SUG1*) and *CIM5* are 26S protease subunits. a, Specificity of anti-*CIM* antibodies. Immunoblots of whole yeast cell extracts (20 μ g of total protein) reacted with affinity-purified anti-*CIM3* and anti-*CIM5* polyclonal antibodies at a dilution of 1/5,000 (20 ng ml⁻¹ of immunoglobulin). Lanes: 1, Wild-type strain; 2, *cim5::HIS3/pFL44-CIM5* (multicopy *CIM5*); 3, *cim3::HIS3/pFL44-CIM3* (multicopy *CIM3*); 4, *cim5::HIS3/pNV11-MSS1*. The same membranes were re-probed with anti-CDC28 antibodies to normalize to the in-varying quantity of CDC28 in the extracts. (+) p54-CIM5; (●) p49-MSS1; (*) p43-CIM3-(*SUG1*); (\$) p34-CDC28. The anti-*CIM5* antibodies strongly recognize (lane 4) the human *MSS1* protein. b, Apparent homologues of *CIM3* and *CIM5* are components of the *Drosophila* 26S protease. Middle two lanes: SDS-PAGE separation of a near-homogeneous preparation of the 26S protease from *D. melanogaster*⁴ stained with Coomassie blue. Proteins in the M_r range 35–120K (marked with circles) are specifically associated with the 26S protease. Proteins of 20–32K (vertical bars) are found in both the 20S and 26S protease particles. Immunoblots of the same preparation were probed with affinity-purified anti-*CIM3* and anti-*CIM5* antibodies (100 ng ml⁻¹). A single cross-reacting species was revealed in each case that comigrated with the 26S protease subunits present in the seventh (anti-*CIM5*) and eighth (anti-*CIM3*) largest bands of the preparation. These bands are composed of two or more comigrating subunits under these gel conditions. The anti-*CIM3* and anti-*CIM5* rabbit polyclonal antibodies were generated using *CIM* protein antigens purified from inclusion bodies after overexpression in *E. coli*²⁷. Immunoblots were developed with the ECL system (Amersham). c, Pulse-chase analysis of Ub-Pro- β gal and Leu- β gal degradation. Exponentially growing wild-type (WT), *cim3-1* and *cim5-1* strains



were labelled with ³⁵S-methionine for 5 min at 30 °C (time 0 lane), followed by a chase in the presence of unlabelled methionine and cycloheximide for 10 and 30 min as indicated. β -galactosidase derivatives were immunoprecipitated and analysed by SDS-PAGE as described²⁸. Radioactivity was quantified with a PhosphorImager (Molecular Dynamics). Autoradiograms show Ub-Pro- β gal is stabilized in the *cim* mutants, whereas Leu- β gal degradation is hardly affected (half-life of 8 min in the *cim* mutants, 7 min in the wild type). Likewise, Arg- β gal remains unstable in the *cim* mutants, with half-lives of 6 min in *cim3-1*, 7.5 min in *cim5-1*, and 6 min in the wild type (data not shown). Vertical lines alongside gels indicate multi-ubiquitinated β -galactosidase derivatives. Multi-ubiquitination of Ub-Pro- β gal was inhibited in the *cim* mutants and in the *ubc4* (ref. 12), *pre1-1* and *pre2-2* mutants^{13,14}. A stable 90K cleavage product of the β -galactosidase derivatives¹⁰ is indicated (*).

FIG. 4 Histone-H1 kinase activity of CDC28 increases with CLB2 and CLB3 protein levels in division-arrested *cim* mutants. **a**, CDC28 protein kinase assays *in vitro* from whole-cell extracts²⁹. Log-phase cultures of the wild type (WT), *cim3-1*, *cim5-1*, *cdc28-4* and *cdc16* were grown at their permissive temperature and then transferred to 37 °C. Samples were withdrawn every hour for 4 h and assayed for CDC28-dependent kinase activity at 37 °C using histone H1 as a substrate. A 3-fold higher activity was observed for *cim3-1* and *cim5-1* after 4 h at 37 °C relative to 28 °C. The kinase activity measured in these assays is CDC28-dependent as the *cdc28-4* mutant had little kinase activity at 24 °C or 37 °C. The *cdc16* mutant arrests division in G2/mitosis³⁰ with high kinase activity. **b**, Immunoblots showing accumulation of CLB2 and CLB3 in *cim* mutants. Levels of CLB2 and CLB3 were compared for *cim* mutants and the isogenic wild type during exponential growth at 30 °C and after α -factor arrest of the mutants at 30 °C, followed by release at 37 °C for the times indicated above lanes (2, 4, 6 or 8 h). The percentage of unbudded cells (including shmoo-shaped cells induced by α -factor treatment) determined for each culture is indicated under each lane. CLB levels are substantially higher in *cim* mutants at 30 °C than in wild type. When cells were treated with α -factor to induce G1 arrest, several unusual effects were noted. Even after extended incubation of the *cim* mutants with α -factor at the permissive temperature, a fraction of the cells remained arrested, with large buds and a residual level of CLB2 and CLB3 expression. When



α -factor was removed from the *cim* mutants at 30 °C or 37 °C, only 50–70% of the cells were able to recover and resume cell division, although this was slower than in wild type. These results suggest that the 26S protease functions in the adaptation of cells to α -factor. The isogenic wild type showed a normal response to α -factor, with the expected fluctuations¹⁵ of CLB2 and CLB3 upon release from arrest at 37 °C (not shown). The immunoblot analysis using whole-cell extracts and affinity-purified anti-CLB2 and anti-CLB3 antibodies was done as previously described¹⁵. Membranes were reprobed with an anti-CDC28 antibody in order to normalize the levels of CDC28 protein in the extracts. The lane marked *clbΔ* shows extracts from strains deleted for either CLB2 or CLB3, whereas *GAL-CLB* indicates strains in which CLB2 or CLB3 was overexpressed from the GAL promoter¹⁵.

required for chromosome segregation. The co-lethality between *cdc28-1N* and the *cim* mutants would be explained if the function of the ubiquitin proteolytic system involved in exiting mitosis were activated by CDC28 (ref. 16). Further investigation will establish whether G2/metaphase arrest in *cim* mutants is due to their inability to degrade an ensemble of cyclins and/or other proteins whose destruction might be necessary before anaphase can occur^{18,20}. □

stabilize Ub-Pro- β gal, Leu- β gal and Arg- β gal^{13,14}.

Histone H1 kinase activity associated with CDC28 increases in the *cim3-1* and *cim5-1* cells arrested in G2/metaphase (Fig. 4a). The active form of the CDC28 kinase in G2/metaphase is composed of the CDC28 catalytic subunit complexed with a mitotic cyclin¹⁵. Degradation of mitotic cyclins occurs *in vitro* through a ubiquitin-dependent pathway^{16,17}, so we examined levels of two B-type cyclins in the *cim* mutants. Immunoblot analysis showed that CLB2 and CLB3 cyclins were already present in greater than normal amounts in the *cim* mutants at 30 °C compared to wild-type (Fig. 4b). This steady-state accumulation of CLB2 and CLB3 is correlated with slower doubling times and a higher fraction of cells in G2/metaphase for the *cim* mutants at 24 °C or 30 °C compared with wild type (Fig. 1, and data not shown). When *cim* mutants were treated with α -pheromone to induce G1 arrest at 30 °C, levels of CLB2 and CLB3 were greatly diminished (Fig. 4b). After release from α -factor arrest at 37 °C, CLB2 and CLB3 accumulated at the time that the *cim* mutants underwent division arrest (Fig. 4b).

Inhibition of ubiquitin-mediated protein degradation prevented chromosome segregation in *Xenopus* egg extracts¹⁸. Our finding that the *cim3-1* and *cim5-1* 26S protease mutants arrest specifically in G2/metaphase, together with the results of Gordon *et al.* that inactivation of a distinct 26S protease subunit of *Schizosaccharomyces pombe* leads to a metaphase arrest¹⁹, provides the first *in vivo* evidence that the 26S protease is

Received 18 May; accepted 12 October 1993.

- Swaffield, J. C., Bromberg, J. F. & Johnston, S. A. *Nature* **357**, 698–700 (1992).
- Dubiel, W., Ferrell, K., Pratt, G. & Rechsteiner, M. *J. biol. Chem.* **267**, 22699–22702 (1992).
- Shibuya, H. *et al. Nature* **357**, 700–702 (1992).
- Udvardy, A. J. *biol. Chem.* **268**, 9055–9062 (1993).
- Forsburg, S. L. & Nurse, P. A. *Rev. Cell Biol.* **7**, 227–256 (1991).
- Piggott, J. R., Ral, R. & Carter, B. L. *A. Nature* **298**, 391–393 (1982).
- Surana, U. *et al. Cell* **65**, 145–161 (1991).
- Byers, B. & Goetsch, L. *Cold Spring Harbor Symp. Quant. Biol.* **38**, 123–131 (1973).
- Hershko, A. & Ciechanover, A. *A. Rev. Biochem.* **61**, 761–807 (1992).
- Bachmair, A., Finley, D. & Varshavsky, A. *Science* **234**, 179–186 (1986).
- Varshavsky, A. *Cell* **69**, 725–735 (1992).
- Johnson, E. S., Bartel, B., Seufert, W. & Varshavsky, A. *EMBO J.* **11**, 497–505 (1992).
- Richter-Rudolf, B., Heinemeyer, W. & Wolf, D. H. *FEBS Lett.* **302**, 192–196 (1992).
- Seufert, W. & Jentsch, S. *EMBO J.* **11**, 3077–3080 (1992).
- Grandin, N. & Reed, S. I. *Molec. cell. Biol.* **13**, 2113–2125 (1993).
- Glutzer, M., Murray, A. W. & Kirschner, M. W. *Nature* **349**, 132–138 (1991).
- Hershko, A., Ganoth, D., Pehrson, J., Palazzo, R. E. & Cohen, L. H. *J. biol. Chem.* **266**, 16376–16379 (1991).
- Holloway, S. L., Glotzer, M., King, R. W. & Murray, A. W. *Cell* **73**, 1393–1402 (1993).
- Gordon, C., McGurk, G., Dillon, P., Rosen, C. & Hastie, N. D. *Nature* **366**, 355–357 (1993).
- Surana, U. *et al. EMBO J.* **12**, 1969–1978 (1993).
- Hutter, K.-J. & Eipel, H. E. *J. gen. Microbiol.* **113**, 369–375 (1979).

22. Pringle, J. R., Adams, A. E. M., Drubin, D. & Haarer, B. K. *Meth. Enzym.* **194**, 565–602 (1991).
23. Gorbatenya, A. E. & Koonin, E. V. *Nucleic Acids Res.* **17**, 8413–8440 (1989).
24. Ohana, B. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 138–142 (1993).
25. Frohlich, K.-U. et al. *J. Cell Biol.* **114**, 443–453 (1991).
26. Boeke, J. D., LaCrute, F. & Fink, G. R. *Molec. gen. Genet.* **197**, 345–346 (1984).
27. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. J. *Meth. Enzym.* **185**, 60–89 (1990).
28. Bartel, B., Wünnig, I. & Varshavsky, A. *EMBO J.* **9**, 3179–3189 (1990).
29. Moreno, S., Hayles, J. & Nurse, P. *Cell* **58**, 361–372 (1989).
30. Hartwell, L. H. & Smith, D. *Genetics* **110**, 381–395 (1985).

ACKNOWLEDGEMENTS. M.G. was supported by an EMBO postdoctoral fellowship. We thank N. Grandin for anti-CLB antibodies, M. Blondel and P. Zarov for their contributions to the initial strain construction, S. Brown for assistance with flow cytometry, F. Lacroute for the pFL44 DNA library, K. Matsumoto for the pNV11-MSS1 plasmid, and C. Gordon for discussion. M.G. thanks K. Madura, E. Johnson and A. Varshavsky for plasmids expressing different β -galactosidase derivatives and for discussion. Pulse-chase experiments were carried out in the Varshavsky laboratory.

Role for DNA methylation in genomic imprinting

En Li*, Caroline Beard & Rudolf Jaenisch†

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

THE paternal and maternal genomes are not equivalent and both are required for mammalian development^{1,2}. The difference between the parental genomes is believed to be due to gamete-specific differential modification, a process known as genomic imprinting. The study of transgene methylation has shown that methylation patterns can be inherited in a parent-of-origin-specific manner^{3–7}, suggesting that DNA methylation may play a role in genomic imprinting. The functional significance of DNA methylation in genomic imprinting was strengthened by the recent finding that CpG islands (or sites) in three imprinted genes, *H19*, insulin-like growth factor 2 (*Igf-2*), and *Igf-2* receptor (*Igf-2r*), are differentially methylated depending on their parental origin^{8–12}. We have examined the expression of these three imprinted genes in mutant mice that are deficient in DNA methyltransferase activity¹³. We report here that expression of all three genes was affected in mutant embryos: the normally silent paternal allele of the *H19* gene was activated, whereas the normally active paternal allele of the *Igf-2* gene and the active maternal allele of the *Igf-2r* gene were repressed. Our results demonstrate that a normal level of DNA methylation is required for controlling differential expression of the paternal and maternal alleles of imprinted genes.

Embryos homozygous for a DNA methyltransferase (MTase) mutation die at embryonic day 11 postcoitum (E11) with their genomic DNA substantially demethylated¹³. We used mutant embryos to examine the effect of reduced genomic methyl cytosine levels on the expression of the *H19*, *Igf-2* and *Igf-2r* genes. Analysis of RNA from mutant embryos indicated that the expression of all three genes was altered. The paternal allele of the *H19* gene is silent and the maternal allele is expressed in normal embryos¹⁴. To assess the effect of a reduced level of DNA methylation on *H19* expression, RNA was isolated from normal and MTase mutant embryos at E10.5 and analysed by northern blot hybridization. Expression of the *H19* gene was significantly elevated in all four homozygous mutant embryos (Fig. 1a, lanes 2, 3, 4 and 7) when compared to wild-type and heterozygous embryos (Fig. 1a, lanes 1, 5 and 6). To determine whether this increased expression in mutant embryos was due to the activation of the silent paternal allele, embryos with distinguishable paternal and maternal alleles were obtained from interspecific

crosses between 129/Sv mice and *Mus castaneus* (*M. cas.*) in the MTase-mutant background. In an RNase protection assay, the paternal *M. castaneus* allele was expressed at roughly equal levels as the 129 maternal allele in homozygous embryos (Fig. 1b, lane 3), whereas only the 129 maternal allele was expressed in the heterozygous embryo (Fig. 1b, lane 2). These results indicate that DNA methylation is essential for maintaining the transcriptionally inactive state of the paternal *H19* allele.

The *Igf-2* gene is closely linked to the *H19* gene but reciprocally imprinted, that is, only expressed from the paternal allele¹⁵. To examine *Igf-2* expression in MTase-deficient embryos, RNA

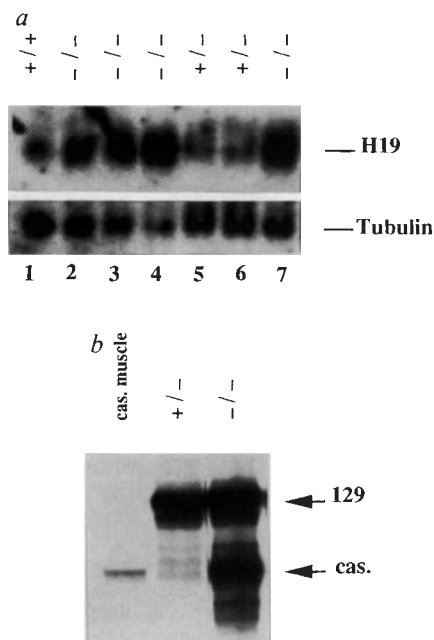
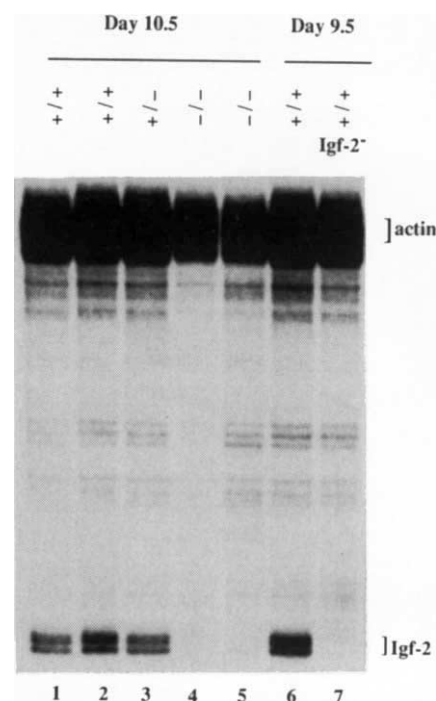


FIG. 1 Expression of *H19* gene in DNA MTase mutant embryos. a, Northern blot analysis of total RNA (10 μ g per lane) from E10.5 embryos hybridized with a 3-kb genomic *EcoRI*–*Sall* fragment which spans the mouse *H19* gene¹⁴ and a tubulin cDNA probe as an RNA loading control. Embryos were wild-type (lane 1) and homozygous (lanes 2, 3, 4 and 7) or heterozygous (lanes 5 and 6) for the *MTase*ⁿ mutation. b, RNase protection assay of total RNA (2 μ g per lane) prepared from E10.5 embryos either homozygous (–/–) or heterozygous (+/–) for the *MTase*ⁿ mutation using a riboprobe generated from a clone containing the 754-bp *Bam*HI–*Stu*I genomic DNA fragment of the *H19* gene as described¹⁴. Only the allele-specific protected fragment (129 or cas., with arrows) from exon 5 is shown. Embryos were obtained by mating F₁ (*M. cas.*/129) male mice heterozygous for the *MTase*ⁿ mutation to 129 female *MTase*ⁿ heterozygous mice, and genotyped for the presence of an *M. cas.* allele of the *H19* gene and the *MTase*ⁿ mutation. Only embryos that carry a paternally inherited *M. cas.* allele and a maternally inherited 129 allele of the *H19* gene were used for RNase protection. An RNA sample from adult *M. cas.* muscle, where *H19* is expressed at a lower level than in embryos, was included as a positive control. Similar results were obtained from 4 out of 4 homozygous mutant embryos. METHODS. E10.5 embryos were dissected in ice-cold saline. DNA was prepared from yolk sac membranes and genotyped for the *MTase*ⁿ mutation by Southern blot analysis as described¹³. To discriminate between *H19* alleles arising from a *M. domesticus* and *M. castaneus* cross, DNA was digested with *Pvu*II and subjected to Southern blot hybridization using the 3-kb *EcoRI*–*Sall* genomic fragment of the *H19* gene¹⁴ (data not shown). RNA was prepared from embryos by the guanidium isothiocyanate/phenol/chloroform method²⁰. RNA samples were treated with 20 units ml^{–1} RNase-free DNase (BRL) for 45 min at 37°C, phenol-extracted, and ethanol precipitated before use. Total RNA (10 μ g) was analysed on a formaldehyde gel and hybridized with the 3-kb *EcoRI*–*Sall* fragment, and the filter was washed at 65°C with 0.2 $\times$ SSC, 0.2% SDS. The filter was stripped and hybridized with a probe derived from tubulin. RNase protection was as described¹⁴.

* Present address: Cardiovascular Research Center, Massachusetts General Hospital-East, 149 13th Street, Charlestown, Massachusetts 02129, USA.

† To whom correspondence should be addressed.

FIG. 2 RNase protection analysis of *Igf-2* gene expression. Total cell RNA was prepared from pools of embryos (2–5 embryos per pool) as described in the legend to Fig. 1. Samples of RNA (10 µg per lane) from E10.5 embryos homozygous for the *MTase*ⁿ mutation (lanes 4 and 5), wild-type (lanes 1 and 2) and heterozygous (lane 3) embryos, and E9.5 wild-type control embryos (lanes 6 and 7) were analysed by RNase protection using the E2 probe containing exon 2 of the *Igf-2* gene¹⁵ and the mouse cytoplasmic actin probe as a control. The assay was as described¹⁵. RNA from an E9.5 embryo carrying a paternally inherited mutant allele of the *Igf-2* gene (*Igf-2*⁻, lane 7) represents a negative control for *Igf-2* expression as exon 2 is deleted in the mutant paternal allele and the maternal allele is transcriptionally silent¹⁵. Because the development of *MTase* mutant embryos is delayed by about one day when compared to normal littermates¹³, RNA from E9.5 wild-type embryos was also analysed.



was isolated from E10.5 embryos and analysed by RNase protection¹⁵. Although the *Igf-2* gene was expressed in wild-type and heterozygous embryos at E10.5 (Fig. 2, lanes 1–3), its expression was almost undetectable in the homozygous embryos (lane 4 and 5). Wild-type E9.5 embryos express *Igf-2* at a similar level to E10.5 embryos (lane 6). Our results therefore suggest that a normal level of DNA methylation is required for the expression of the active paternal allele.

The *Igf-2r* gene, located on chromosome 17, is expressed exclusively from the maternal allele¹⁶. To determine whether *Igf-2r* expression was also affected by a reduction in DNA methylation, RNA was isolated from E9.5 embryos and analysed by RNase protection. *Igf-2r* expression was almost the same in

MTase mutant embryos (Fig. 3, lane 2) as in the wild-type embryos (Fig. 3, lane 1). These results suggest that the maintenance of the *Igf-2r* gene imprint might not involve DNA methylation. Alternatively, it seemed possible that the maintenance of the *Igf-2r* gene imprint, although dependent on DNA methylation, was quantitatively less affected by demethylation of genomic DNA than that of the other two imprinted genes. In this context, note that the *MTase* mutation, generated by insertion of a targeting vector into the *NaeI* site at the 5' end of the *MTase* gene¹³ (termed *MTase*ⁿ allele), is a partial loss of function mutation. To address whether the level of DNA methylation was critical for the maintenance of the *Igf-2r* imprint, we used a second mutation which was generated by disrupting the *MTase*

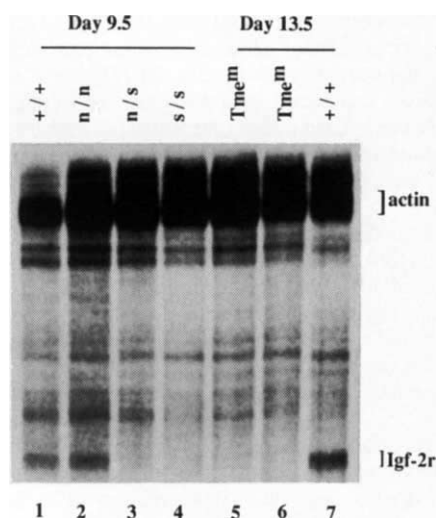


FIG. 3 RNase protection analysis of *Igf-2r* expression. Embryos were genotyped by Southern blot analysis of DNA from half of the yolk sac membranes using pBB (for *MTase*ⁿ)¹³ or pHH (for *MTase*^s) probes. Because the level of *Igf-2r* expression is very low in E9.5 embryos, embryos plus the other half of the yolk sac membranes of identical genotype were pooled (3–4 embryos per pool) for RNA preparation using the method described in the legend to Fig. 1. Samples of RNA (10 µg per lane) from E9.5 wild-type (lane 1), *MTase*ⁿ homozygotes (lane 2, *n/n*), compound heterozygotes *MTase*ⁿ/*MTase*^s (lane 3, *n/s*) and *MTase*^s homozygotes (lane 4, *s/s*), and E13.5 *Tme*^m (lane 5 and 6) and wild-type (lane 7) embryos (RNA kindly provided by Denise Barlow) were analysed by RNase protection. *Tme*^m stands for maternally inherited *Tme* mutation which contains a large chromosomal deletion spanning the *Igf-2r* gene. RNA from *Tme*^m embryos serves as a negative control because the maternally inherited *Igf-2r* allele is deleted while the paternal allele is silent¹⁶. The probe was prepared from a plasmid SN that contains a 347-bp *SmaI*–*NotI* fragment spanning the transcription start sites and part of the first exon of the *Igf-2r* gene¹². A 190-base antisense riboprobe was synthesized *in vitro* from a *HinfI* linearized DNA template using T3 RNA polymerase and [³²P]CTP. The size of the protected fragment was 155 bases. RNase protection was as described²¹ except that hybridization was done at 62°C and RNase digestion was at 37°C. Mouse cytoplasmic actin probe was added as a quantitative control. The protected *Igf-2r* fragment and actin fragment are indicated. Because the development of the *MTase*^s/*MTase*^s mutant embryos is delayed by about 1 day compared to wild-type embryos, RNA isolated from wild-type E8.5 embryos was analysed as a control and found to contain a similar level of expression as in E9.5 embryos (data not shown).

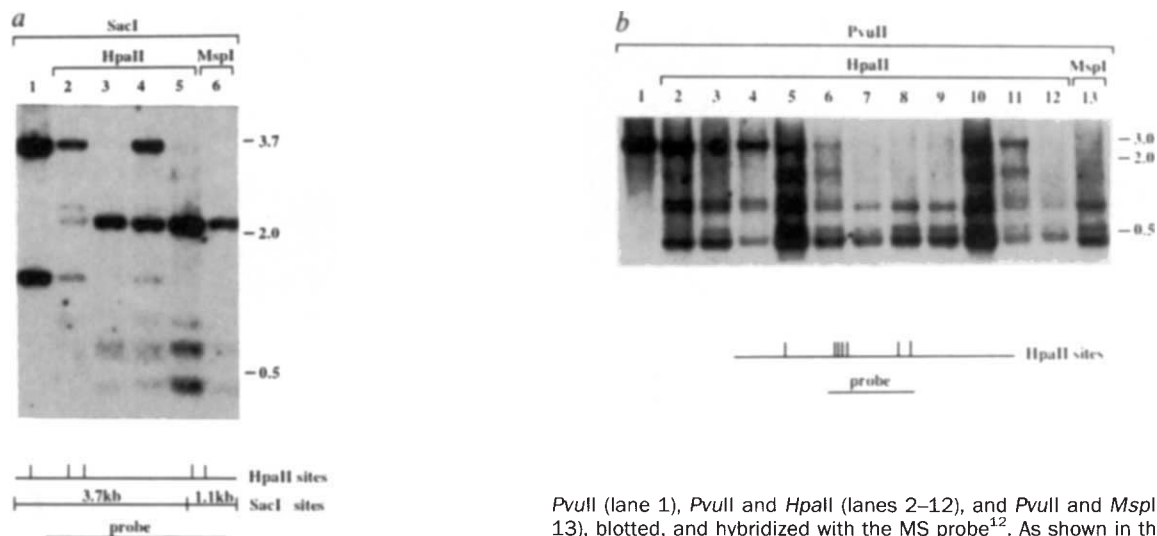


FIG. 4 a, Methylation of the region 5' to the *H19* gene in mutant embryos and ES cells. Genomic DNA was digested with *SacI* (lane 1), *SacI* and *HpaII* (lanes 2–5), or *SacI* and *MspI* (lane 6), blotted, and hybridized with the 4.0 R1 probe⁹. The *HpaII* sites and *SacI* sites of this region, and the probe are diagrammed at the bottom. Two *SacI* fragments, 3.7 kb and 1.1 kb, were detected using the 4.0 R1 probe (lane 1). Digestion of genomic DNA with *SacI* and *HpaII* showed that *SacI* fragments were partially sensitive to *HpaII* in wild-type J1 cells (lane 2) and a *MTase*^{+/+} heterozygous embryo (lane 4), consistent with the maternal allele being unmethylated, whereas the *SacI* fragments were almost completely digested by *HpaII* in homozygous ES cells, line 52, and a *MTase*⁰ homozygous mutant (lanes 3 and 5). DNA size markers (kb) shown on the right. b, Methylation of the CpG island in region 2 of the *Igf-2r* gene in mutant embryos and ES cells. DNA was digested with

PvuII (lane 1), *PvuII* and *HpaII* (lanes 2–12), and *PvuII* and *MspI* (lane 13), blotted, and hybridized with the MS probe¹². As shown in the diagram at the bottom, the 3-kb *PvuII* fragment contains seven *HpaII* sites. When digested with *PvuII* and *HpaII*, the *PvuII* fragment of wild-type J1 ES cells (lane 2) and a heterozygous embryo (lane 4) was partially digested by *HpaII* due to the presence of the unmethylated paternal allele, whereas the *PvuII* fragment of the homozygous ES line 52 (lane 3) and mutant embryos (lanes 5–12) was much less resistant to *HpaII* digestion, indicating loss of methylation on the maternal allele. A substantial amount of the *PvuII* fragment was still present in the *MTase*⁰/*MTase*⁰ embryos (lanes 5, 6, 10 and 11), but almost undetectable in the *MTase*⁰/*MTase*^s (lanes 7–9, 12) embryos, consistent with the *MTase*^s being a more severe mutation. The substantial difference in DNA methylation between *MTase*⁰/*MTase*⁰ and *MTase*⁰/*MTase*^s embryos might account for the difference in *Igf-2r* expression (see Fig. 3). Embryonic DNA was prepared and analysed by Southern blot hybridization using the methods as described previously¹³ with some modification. DNA size markers (kb) shown on the right.

gene further downstream at a *SalI* site (termed *MTase*^s) corresponding to amino-acid residue 310. The *MTase*^s allele is a more severe mutation than the *MTase*⁰ allele as *MTase*^s/*MTase*⁰ embryos die at the 5–10 somite stage with their genomic DNA more demethylated than *MTase*⁰/*MTase*⁰ embryos (E.L. *et al.*, manuscript in preparation). RNA was isolated from E9.5 *MTase*⁰/*MTase*^s compound heterozygous and *MTase*⁰/*MTase*⁰ homozygous mutant embryos and analysed by RNase protection. The *Igf-2r* gene was repressed in both the *MTase*⁰/*MTase*^s and the *MTase*^s/*MTase*^s mutant embryos (Fig. 3, lanes 3 and 4), in contrast to the *MTase*⁰/*MTase*⁰ embryos. These results indicate that DNA methylation is required for expression of the maternal *Igf-2r* allele. *Igf-2r* gene expression appears, however, to be quantitatively less sensitive to genomic demethylation than *H19* and *Igf-2* gene expression.

We have examined the methylation patterns of the *H19* (Fig. 4a) and *Igf-2r* (Fig. 4b) genomic regions using methylation-sensitive *HpaII* digestion followed by Southern blot analysis. The differentially methylated region of the *H19* gene^{8–10} was significantly demethylated in *MTase*⁰/*MTase*⁰ mutant embryos as well as in homozygous ES cells (lanes 3, 5; see also *MspI* digest in lane 6) as compared to wild-type ES cells or normal embryos (lanes 2, 4), suggesting that the methylation difference between the parental alleles was erased in mutant ES cells and embryos. The CpG island in region 2 of the *Igf-2r* gene, which is differentially methylated¹², was also substantially demethylated in *MTase*⁰ homozygous mutant embryos (Fig. 4b, lanes 5, 6, 10, 11) as compared to control ES cells or embryos (lanes 2, 4), but to a lesser extent than that of the *H19* gene, suggesting that in mutant embryos methylation of this CpG island might be preferentially maintained as compared to the 5' region of the *H19* gene. The level of island methylation was, however, significantly lower in the more severe *MTase*⁰/*MTase*^s compound heterozy-

gous than in *MTase*⁰ homozygous mutant embryos (lanes 7, 8, 9, 12; also compare *MspI* digest in lane 13). This is consistent with normal *Igf-2r* expression in *MTase*⁰/*MTase*⁰ versus lack of expression in *MTase*⁰/*MTase*^s mutant embryos and suggests that the island must be substantially demethylated to affect expression of the gene. Our results imply that DNA methylation is causally involved in maintaining allele-specific expression of imprinted genes.

Although numerous experiments have correlated DNA methylation with gene repression and demethylation with gene activation^{17,18}, experimental proof that DNA methylation controls the activity of cellular genes *in vivo* has been lacking. The activation of the normally silent paternal *H19* allele in *MTase*-deficient embryos provides first *in vivo* evidence for a causal link between DNA methylation and gene activity. Interference with the maintenance of CpG island methylation in the *H19* promoter region^{8–10} may lead to loss of the imprint and activation of the gene.

In contrast to the activation of the *H19* gene, it is less clear how genomic DNA demethylation could lead to the inactivation of the expressed *Igf-2* and *Igf-2r* alleles. It has been proposed that expression of the two reciprocally imprinted genes *H19* and *Igf-2* is functionally and/or mechanistically related and that the imprinting of a single chromosomal site might control the activity of both genes⁹. Our results are consistent with the possibility that *H19* is the 'primary' imprinted gene whose activity suppresses *in cis* the expression of the *Igf-2* gene. Because allele-specific methylation has been detected in both *Igf-2* and *Igf-2r* genes^{10–12}, it is also possible that expression of both genes is directly controlled by DNA methylation. It has recently been proposed that the CpG island in the intron of the *Igf-2r* gene represents the 'imprinting box' that acts as a 'gene silencer' by binding to a repressor protein when not methylated¹². Interfer-

ence with methylation of the box might inhibit expression.

In conclusion, we have demonstrated that DNA methylation plays a critical role in genomic imprinting. Specifically, DNA methylation is required for the maintenance of monoallelic expression of imprinted genes. Our data do not address, however, whether DNA methylation plays a functional role in establishing imprinting patterns in the male and female gametes. Although it has been shown that the maternally inherited *Igf-2r* gene and the TG-A transgene acquire their distinctive methylation patterns during oogenesis^{12,19}, the molecular mechanism of genomic imprinting in developing gametes is still largely unknown and is a challenging issue. □

Received 4 June; accepted 16 September 1993.

1. Surani, M. A. H., Barton, S. C. & Norris, M. L. *Nature* **306**, 548–550 (1984).
2. McGrath, J. & Solter, D. *Cell* **37**, 179–183 (1984).
3. Reik, W., Collick, A., Norris, M. L., Barton, S. C. & Surani, M. A. *Nature* **328**, 248–251 (1987).
4. Sapienza, C., Paterson, A. C., Rossant, J. & Balling, R. *Nature* **328**, 251–254 (1987).

5. Swain, J. L., Stewart, T. A. & Leder, P. *Cell* **50**, 719–727 (1987).
6. Hadchouel, M., Farza, H., Simon, D., Tiollais, P. & Pourcel C. *Nature* **329**, 454–456 (1987).
7. Sasaki, H. J. et al. *Development* **111**, 573–581 (1991).
8. Ferguson-Smith, A. C., Sasaki, H. J., Cattanach, B. M. & Surani, M. A. *Nature* **362**, 751–755 (1993).
9. Bartolomei, M. S. et al. *Genes Dev.* **7**, 1663–1673 (1993).
10. Brandeis, M. et al. *EMBO J.* **12**, 3669–3677 (1993).
11. Sasaki, H. et al. *Genes Dev.* **6**, 1843–1856 (1992).
12. Stöger, R. et al. *Cell* **73**, 61–71 (1993).
13. Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915–926 (1992).
14. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. *Nature* **351**, 153–155 (1991).
15. DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. *Cell* **64**, 849–859 (1991).
16. Barlow, D. P., Stöger, R., Herrmann, B. G., Saito, K. & Schweifer, N. *Nature* **349**, 84–87 (1991).
17. Cedar, H. *Cell* **53**, 3–4 (1988).
18. Bird, A. P. *Nature* **321**, 209–213 (1986).
19. Chaillet, J. R., Vogt, T. F., Beier, D. R. & Leder, P. *Cell* **66**, 77–83 (1991).
20. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
21. Melton, D. A. et al. *Nucleic Acids Res.* **12**, 7035–7056 (1984).

ACKNOWLEDGEMENTS. We thank A. Efstratiadis for the *Igf-2* knock-out mice and *Igf-2* probes, D. P. Barlow for *Igf-2r* probes and *Tme*⁺ RNA, S. M. Tilghman for *H19* probes, D. P. Barlow and M. S. Bartolomei for helpful discussion and B. Bates, L. Jackson-Grusby, R. Jütermann, J. Kreidberg, P. Soloway and W. Strauss for critical reading of the manuscript. This work was supported by a grant from the NIH.

Yeast excision repair gene *RAD2* encodes a single-stranded DNA endonuclease

Yvette Habraken^{*†}, Patrick Sung^{†‡},
Louise Prakash^{*†} & Satya Prakash^{†§}

^{*}Department of Biophysics and [†]Department of Biology, University of Rochester, Rochester, New York 14642, USA

[†]Present addresses: UTMB, Sealy Center for Molecular Science, Medical Research Building, Route J61, 11th Street and Mechanic, Galveston, Texas 77555-1061, USA.

[§]To whom correspondence should be addressed at UTMB.

IN eukaryotes nucleotide excision repair of DNA damaged by ultraviolet radiation requires several gene products; defects in this process result in the cancer-prone syndrome xeroderma pigmentosum (XP) in humans^{1,2}. The *RAD2* gene is one of at least seven genes indispensable for excision repair in the yeast *Saccharomyces cerevisiae*², and its encoded protein shares remarkable homology with the XP group-G gene product³. Here we overproduce the *RAD2*-encoded protein in *S. cerevisiae*, purify it to near homogeneity, and show that RAD2 protein in the presence of magnesium degrades circular single-stranded DNA. The RAD2 endonuclease is specific for single-stranded DNA as it does not act on double-stranded DNA. Given the absolute requirement for *RAD2* in the incision step of excision repair, our findings directly implicate RAD2 protein and its human homologue XPG protein as a catalytic component that incises the damaged DNA strand during excision repair. Furthermore, our results indicate that eukaryotes probably employ two distinct endonuclease activities to mediate the dual incision at the damage site.

RAD2 protein was overproduced in yeast (Fig. 1a) using the vector pPE280 (ref. 4), and purified ~3,000-fold to near homogeneity (Fig. 1b). Purified RAD2 protein migrates in denaturing

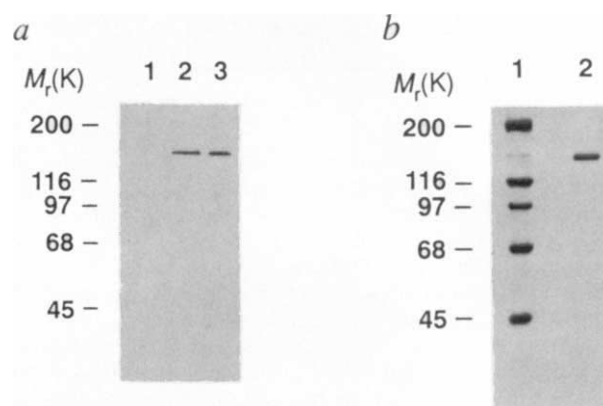
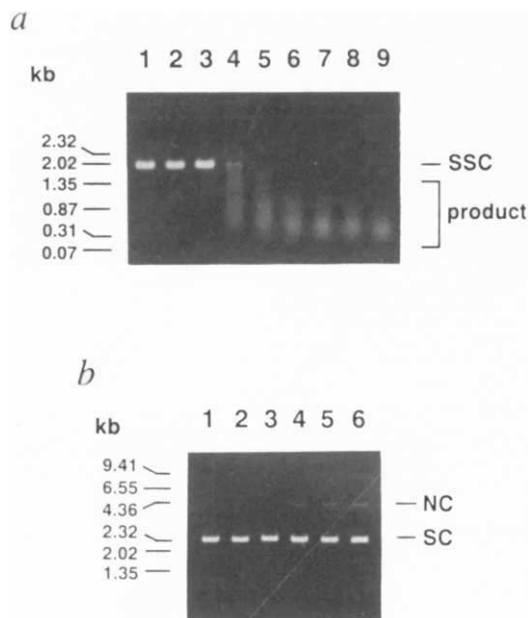


FIG. 1 Overproduction and purification of RAD2 protein. a, Nitrocellulose blot of a 7% denaturing polyacrylamide gel was probed with affinity-purified anti-RAD2 antibodies. Lane 1, cell extract from yeast strain LY2 harbouring the vector pPE280 (GAL/PGK), which lacks the *RAD2* gene; lane 2, extract from LY2 strain harbouring the *RAD2*-overproducing plasmid pR2.26 (GAL/PGK-RAD2); lane 3, 10 ng purified RAD2. b, Coomassie-blue staining of a denaturing polyacrylamide gel. Lane 1, molecular weight standards; lane 2, 1 µg purified RAD2.

METHODS. Polyclonal antibodies were raised in rabbits using a β -galactosidase–RAD2 fusion protein produced in *E. coli* carrying the plasmid pKM8 (ref. 6). This insoluble fusion protein consists of the first 26 amino acids of β -galactosidase followed by the last 861 amino acids of RAD2.

For overproduction of full-length RAD2 protein in yeast, the *RAD2* gene⁵ (from position –24 to 53 nucleotides downstream of the TAA stop codon) was placed under the control of the synthetic galactose-inducible hybrid GAL/PGK promoter in the vector pPE280 (ref. 4), resulting in plasmid pR2.26 (GAL/PGK-RAD2). The protease-deficient yeast strain LY2 (*MATa leu2-3 leu2-112 gal1 trp-1Δ reg1-501 pep4-3 prb112*) was used as host for pR2.26. Yeast cultures (1 litre each) were grown overnight to stationary phase in synthetic medium lacking leucine, diluted ten times with YPD medium containing 1% galactose, and incubated in fermentors at 30 °C until the titre reached 7×10^7 cells ml⁻¹. All subsequent steps were carried out at 4 °C. Cell lysates were prepared from yeast cells (150 g) passed through a French press¹²; the clear supernatant from centrifugation at 100,000g for 90 min was treated with ammonium sulphate at 0.21 g ml⁻¹ and the precipitated proteins were pelleted (20,000g, 30 min), dissolved in 100 ml buffer K (20 mM KH₂PO₄, pH 7.4, 10 mM β -mercaptoethanol, 0.5 mM EDTA, 10% glycerol) containing protease inhibitors¹² and dialysed overnight against 3 litres of the same buffer. The dialysate (fraction 1) was applied to a Q-Sepharose column (2.5 × 16 cm), which was eluted with a 550-ml gradient of 0–400 mM KCl. Fractions containing RAD2 protein eluted at ~300 mM, as confirmed by immunoblotting, and were pooled (fraction 2, 42 ml) and treated with ammonium sulphate at 0.24 g ml⁻¹. Precipitated proteins were pelleted (20,000g for 30 min), dissolved in 30 ml K, and dialysed for 3 h against buffer K to lower the ionic strength to ~50 mM KCl. Proteins in the dialysate were fractionated in a column of S-Sepharose (1 × 10 cm) with an 80-ml gradient of 0–360 mM KCl. The pool of RAD2 protein (fraction 3, 8 ml) eluting at ~180 mM KCl was concentrated to 0.8 ml using Centricon-30 (Amicon) and subjected to molecular size exclusion on a Sephacryl S300 column (1 × 40 cm) in buffer K with 40 mM KCl. The RAD2 pool from S300 (fraction 4, 3.5 ml) was purified further in a Mono-S column (HR5/5) using a 30-ml, 0–200 mM KCl gradient. RAD2 protein (fraction 5, about 100 µg in 5 ml) elutes from Mono S at ~150 mM KCl and, after concentration to 500 µg ml⁻¹ or higher in Centricon-30 (Amicon), was stored in small aliquots at –70 °C. Three different preparations of apparently homogeneous RAD2 protein all give comparable results in enzymatic assays.



polyacrylamide gels as a protein of relative molecular mass (M_r) 130,000 (130K) (Fig. 1a, b), which is consistent with the predicted size of 118K (ref. 5). RAD2 is very scarce in wild-type yeast extract and enrichment by immunoprecipitation is necessary to detect it⁶. RAD2 immunoprecipitated from wild-type extract runs as the same size on denaturing polyacrylamide gels as RAD2 purified from the overproducing strain⁶.

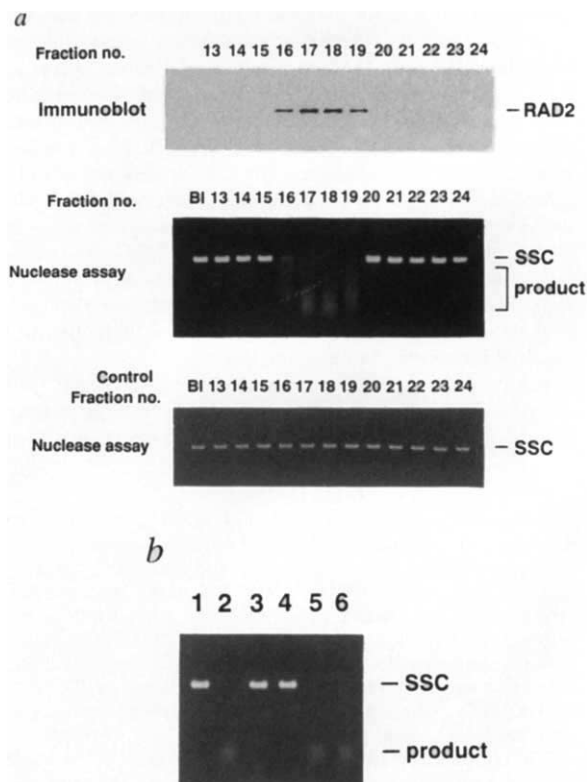
Agarose gel electrophoresis and staining with ethidium bromide show that purified RAD2 protein degrades circular single-stranded DNA (Fig. 2a). As reaction time increases, the single-stranded DNA substrate is converted to forms with higher electrophoretic mobility (Fig. 2a, lanes 4–9). The nuclease activity

is endonucleolytic as it acts on covalently closed single-stranded M13 DNA (Fig. 2a) and Φ X174 DNA (data not shown). Nuclease action is not affected by 1 mM ATP (data not shown) but requires Mg^{2+} (Fig. 2a; compare lanes 2 and 9), which cannot be substituted by Ca^{2+} , Co^{2+} , Mn^{2+} or Zn^{2+} (data not shown). Nuclease activity is maximal at pH 8.0, is inhibited by KCl concentrations ≥ 30 mM (data not shown), and is abolished by 0.5% SDS (Fig. 2a, lane 3). Under conditions optimal for the degradation of single-stranded DNA, undamaged or ultraviolet-damaged double-stranded supercoiled pBR322 DNA (Fig. 2b) or M13 DNA (data not shown) is not nicked by the RAD2 protein. Over a pH range of 5.5 to 9.0 (in increments of 0.5),

METHODS. Covalently closed single-stranded viral M13 mp18 DNA or pBR322 DNA (mixture of supercoiled and nicked circular forms), 200 ng (600 pmol nucleotides), was incubated at 37 °C with 200 ng (1.7 pmol) RAD2 in 10 μ l buffer R (50 mM Tris-HCl, pH 8.0, 20 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), and 100 μ g ml⁻¹ BSA). The reaction was terminated with 1% SDS and the DNA analysed at 4 °C in 0.9% agarose gels run in 20 mM Tris-acetate, pH 7.5, 0.5 mM EDTA, at 80 V for 1.5 h. Gels were treated with ethidium bromide to stain DNA and then photographed. UV irradiation (500 J m⁻²) of DNA in 20 μ l droplets was done with germicidal lamps emitting at 254 nm.

FIG. 3 Nuclease activity is intrinsic to RAD2. **a**, Co-elution of nuclease activity with RAD2. Fractions 13 to 24 (1.25 ml each) from the Mono-S step of RAD2 purification from extract of RAD2-overproducing strain LY2[pR2.26] were concentrated individually to 150 μ l. Top panel, 0.5 μ l of those concentrated fractions was assayed by immunoblotting for RAD2 content; middle panel, the same concentrated fractions (1 μ l) were assayed for ability to degrade single-stranded viral M13 DNA; bottom panel, equivalent fractions derived from extract of control strain LY2 harbouring vector pPE280 and lacking RAD2 had no nuclease activity. BI, DNA incubated without any column fraction. **b**, Specific inhibition of nuclease activity by anti-RAD2 antibodies. Single-stranded viral M13 DNA was incubated alone (lane 1), with RAD2 (lane 2), and with RAD2 in the presence of anti-RAD2 antibodies (2 μ g and 4 μ g in lanes 3 and 4, respectively), anti-RAD1 antibodies (4 μ g in lane 5), or anti-RAD10 antibodies (4 μ g in lane 6). SSC, single-stranded circular DNA; product, RAD2-digested DNA.

METHODS. Extract from 150 g yeast strain LY2, harbouring the RAD2-overproducing plasmid pR2.26, was processed through the Mono-S step as described in Fig. 1 legend, and the RAD2 peak and flanking fractions were concentrated in Centricon-30 and incubated with covalently closed single-stranded viral M13 DNA for 30 min as described in Fig. 2 legend. For the control, extract from 150 g LY2, harbouring vector pPE280 and lacking the RAD2 insert, was processed in the same way and assayed. The wild-type levels of RAD2 present in the control fractions were too low to detect by immunoblotting (data not shown) or by nuclease assay (bottom panel in a). In the antibody inhibition experiment shown in b, 200 ng purified RAD2 protein was preincubated for 15 min on ice in buffer R (Fig. 2 legend) with affinity-purified anti-RAD1, anti-RAD2 and anti-RAD10 antibodies⁶ before addition of DNA and incubation at 37 °C.



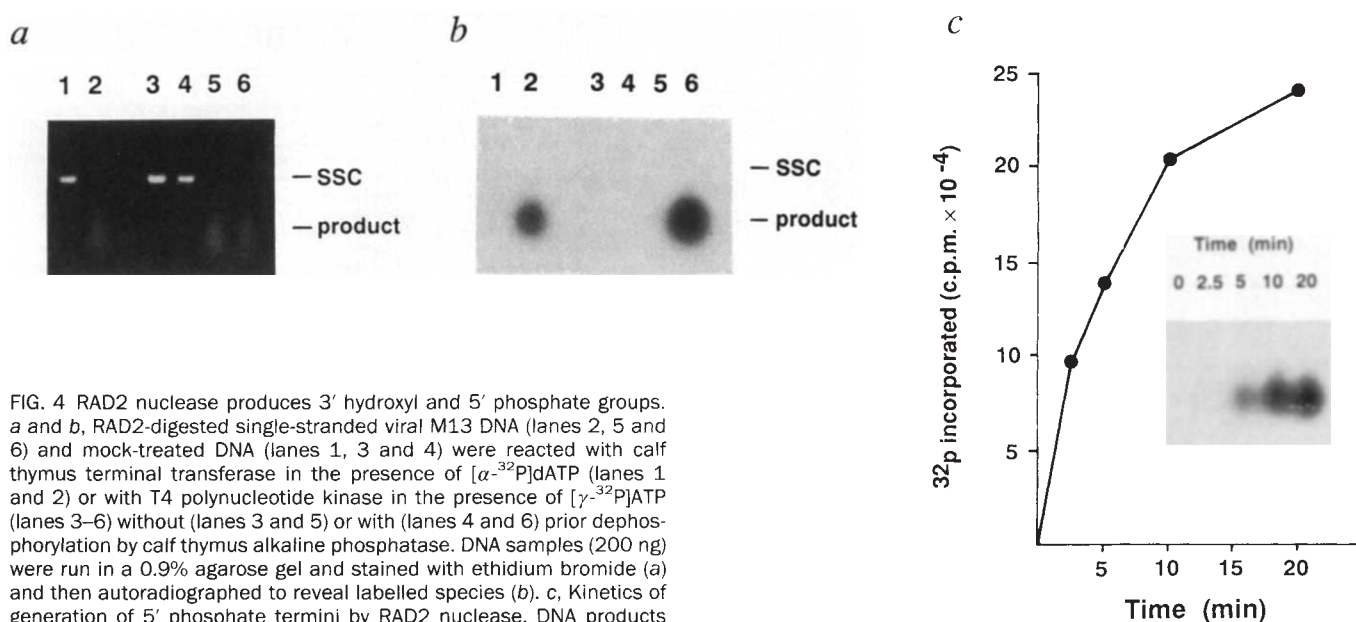


FIG. 4 RAD2 nuclease produces 3' hydroxyl and 5' phosphate groups. a and b, RAD2-digested single-stranded viral M13 DNA (lanes 2, 5 and 6) and mock-treated DNA (lanes 1, 3 and 4) were reacted with calf thymus terminal transferase in the presence of [α - ^{32}P]dATP (lanes 1 and 2) or with T4 polynucleotide kinase in the presence of [γ - ^{32}P]ATP (lanes 3–6) without (lanes 3 and 5) or with (lanes 4 and 6) prior dephosphorylation by calf thymus alkaline phosphatase. DNA samples (200 ng) were run in a 0.9% agarose gel and stained with ethidium bromide (a) and then autoradiographed to reveal labelled species (b). c, Kinetics of generation of 5' phosphate termini by RAD2 nuclease. DNA products after increasing times of incubation were 5'-end labelled with T4 polynucleotide kinase after phosphatase treatment. Incorporation of ^{32}P into DNA was quantified by liquid scintillation counting after precipitation with 10% trichloroacetic acid and collection on GF/C filters¹⁶. Points plotted represent average values from three independent experiments. Labelled reaction products from one experiment were electrophoresed in 0.9% agarose gel and autoradiographed (inset). SSC, single-stranded circular DNA; product, RAD2-digested DNA.

METHODS. For 3'-end-labelling, mock-treated and RAD2-digested DNAs, 2 μg each, were incubated with 25 units of calf thymus terminal transferase (Boehringer Mannheim) and 20 μCi of [α - ^{32}P]dATP (3,000 Ci mmol^{-1} ; Amersham) at 37 °C for 30 min. For 5'-end-labelling, 3 μg DNA was first incubated for 60 min at 37 °C with or without 1.5 units of calf thymus alkaline phosphatase (Boehringer Mannheim). After

phenol extraction, DNA was precipitated with ethanol, redissolved and treated at 37 °C for 30 min with 1.6 units T4 polynucleotide kinase (Boehringer Mannheim) and 15 μCi [γ - ^{32}P]ATP (3,000 Ci mmol^{-1} ; Amersham). DNAs from the 3'- and 5'-labelling mixtures were purified by two rounds of phenol extraction and ethanol precipitation before gel electrophoresis and staining with ethidium bromide or autoradiography (a and b). In c, 2 μg single-stranded viral M13 DNA were incubated with 2 μg RAD2 in 120 μl buffer R at 37 °C (see Fig. 2 legend). Aliquots of 20 μl were withdrawn after 0, 2.5, 5, 10 and 20 min, heated at 100 °C for 2 min to inactivate RAD2, and then treated with 1 unit of calf thymus alkaline phosphatase for 30 min at 37 °C. Dephosphorylated DNA was purified and incubated with 2 units of T4 polynucleotide kinase and 10 μCi of [γ - ^{32}P]ATP at 37 °C for 30 min.

RAD2 showed no activity on negatively supercoiled double-stranded DNA with or without ultraviolet irradiation (data not shown).

The high degree of purity of our RAD2 protein preparation (Fig. 1b) argues that nuclease activity is intrinsic to this protein. To establish this firmly, we did the following experiments. First, we tested whether nuclease activity increases with overproduction of RAD2, and whether this activity copurifies with RAD2. As shown in Fig. 3a, the nuclease activity co-elutes with RAD2 from the RAD2-overproducing strain LY2[pR2.26] (top and middle panels): it is absent in the equivalent fractions derived from extract of the control yeast strain LY2[pPE280] (bottom panel). These control fractions contain only wild-type levels of RAD2, which is too low to detect by immunoblotting or by nuclease assay. Second, we tested whether the nuclease activity could be specifically inhibited by anti-RAD2 antibodies. Figure 3b shows that this activity is strongly inhibited by anti-RAD2 antibodies but not by an excess of antibodies directed against RAD1 or RAD10 proteins⁶. Taken together, our results indicate that the nuclease activity is intrinsic to RAD2.

The DNA product of RAD2 nuclease action can be efficiently labelled with [α - ^{32}P]dATP using calf thymus terminal transferase (Fig. 4a and b, lane 2), indicating that a 3' hydroxyl group is present. The incorporation of ^{32}P catalysed by T4 polynucleotide kinase in the presence of [γ - ^{32}P]ATP requires prior dephosphorylation of the digested DNA with calf thymus alkaline phosphatase (Fig. 4a and b, lanes 5 and 6), indicating that there is a 5' phosphate in RAD2-digested DNA. The kinetics of the production of DNA ends by RAD2 nuclease was followed by ^{32}P -labelling of the 5' end with T4 polynucleotide kinase after phosphatase treatment. ^{32}P -labelled reaction products were

either separated on an agarose gel and visualized by autoradiography (Fig. 4c, inset) or precipitated with trichloroacetic acid and counted for radioactivity (Fig. 4c). More DNA ends are generated by RAD2 as incubation time is increased (Fig. 4c), suggesting that digestion is endonucleolytic.

On the basis that RAD2 is a homologue of XPG (ref. 3), we predict that XPG protein will also possess a single-stranded DNA endonuclease activity. In both human and frog cell extracts, DNA containing a pyrimidine dimer is incised twice by the excision repair machinery, with one cut being made at the twenty-third phosphodiester bond 5' to the lesion, and the other at the sixth phosphodiester bond 3' to the lesion, resulting in a 29-nucleotide fragment containing the lesion^{7,8}. Apart from RAD2, the excision repair proteins RAD1 and RAD10 together specify an endonuclease activity in *S. cerevisiae*^{9,10}. The requirement for two distinct endonucleases in nucleotide excision repair in eukaryotes contrasts with the situation in *Escherichia coli*, in which the UvrB and UvrC proteins involved in dual incision of damaged DNA have no endonuclease activity themselves¹¹.

As the RAD1/RAD10 and RAD2 endonucleases are both indispensable for incision of damaged DNA² but do not by themselves act preferentially on ultraviolet-damaged DNA⁹, they may be targeted to the damage site as a result of their interaction with other components of the excision repair machinery, namely, RAD3, RAD4, RAD14 and RAD25 (ref. 2). We have shown that RAD3 has a DNA-unwinding activity¹² and that RAD14 binds specifically to ultraviolet-damaged DNA¹³. The presence of ATPase and DNA helicase motifs in the RAD25-encoded protein^{14,15} suggests that this protein also functions in DNA unwinding during excision repair.

Unlike RAD2, the RAD1/RAD10 enzyme makes endonu-

cleolytic scissions in unwound regions in negatively supercoiled DNA⁹. Binding of the damaged DNA by RAD14 and other damage recognition components, and subsequent localized unwinding of the helix by RAD3 and RAD25, may prepare the damaged strand for incision by the RAD1/RAD10 endonuclease. The nicked and unwound single-stranded damaged region could then be acted upon by RAD2. The size (29 nucleotides) of this fragment resulting from dual incision may reflect the extent of helix unwinding by the multiprotein excision-repair complex. □

Received 2 September; accepted 14 October 1993.

1. Cleaver, J. E. & Kraemer, K. H. in *The Metabolic Basis of Inherited Disease* 6th edn (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2949–2971 (McGraw-Hill, New York, 1989).
2. Prakash, S., Sung, P. & Prakash, L. *Rev. Genet.* **27**, 33–70 (1993).
3. Scherly, D. et al. *Nature* **363**, 182–185 (1993).
4. Kingsman, S. M. et al. *Meth. Enzym.* **152**, 330–341 (1987).
5. Madura, K. & Prakash, S. *J. Bact.* **166**, 914–933 (1986).
6. Bailly, V., Sommers, C. H., Sung, P., Prakash, L. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8273–8277 (1992).
7. Huang, J. C., Svoboda, D. L., Reardon, J. T. & Sancar, A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3664–3668 (1992).
8. Svoboda, D. L., Taylor, J. S., Hearst, J. E. & Sancar, A. *J. biol. Chem.* **268**, 1931–1936 (1993).
9. Sung, P., Reynolds, P., Prakash, L. & Prakash, S. *J. biol. Chem.* **268**, 26391–26399 (1993).
10. Tomkinson, A. E., Bardwell, A. J., Bardwell, L., Tappe, N. J. & Friedberg, E. C. *Nature* **362**, 860–862 (1993).
11. Orren, D. K. & Sancar, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5237–5241 (1989).
12. Sung, P., Prakash, L., Matson, S. W. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8951–8955 (1987).
13. Guzder, S. N., Sung, P., Prakash, L. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **90**, 5433–5437 (1993).
14. Gulyas, K. D. & Donahue, T. F. *Cell* **69**, 1031–1042 (1992).
15. Park, E. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11416–11420 (1992).
16. Sung, P., Prakash, L. & Prakash, S. *Nature* **355**, 743–745 (1992).

ACKNOWLEDGEMENTS. We thank V. Bailly for affinity-purified anti-RAD2 antibodies, R. Sclafani for the protease-deficient yeast strain, and S. Kingsman for vector pPE280. This work was supported by the US National Cancer Institute and the Department of Energy.

ERRATA

c-Jun is essential for normal mouse development and hepatogenesis

Frank Hilberg, Adriano Aguzzi
Norma Howells & Erwin F. Wagner

Nature **365**, 179–181 (1993)

FIGURE 2 of this letter compares the morphological appearance of *c-jun*^{-/-} and *c-jun*^{+/-} embryos at 12.5 days of gestation, and not of *c-jun*^{-/-} and *c-jun*^{-/-} embryos as stated in the opening sentence of the figure legend. □

A role for sialyl Lewis-X/A glycoconjugates in capillary morphogenesis

Mai Nguyen, Naomi A. Strubel
& Joyce Bischoff

Nature **365**, 267–269 (1993)

THE penultimate sentence of the first paragraph of this letter should start “Anti-Lewis-B and anti-A exhibited the highest binding...” and not “Anti-Lewis-B and anti-lewis-A...” as published. □

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as ‘manuscripts’ and ‘value \$5’ to prevent the imposition of import duty and value-added tax.

A wonderful sight to behold?

James M. Jasper

American Ground Zero: The Secret Nuclear War. By Carole Gallagher. MIT Press: 1993. Pp. 422. \$50, £32.25.

Nuclear Energy and the Public. By Joop van der Pligt. Blackwell: 1993. Pp. 193. £35.

Life Under A Cloud: American Anxiety About the Bomb. By Allan M. Winkler. Oxford University Press: 1993. Pp. 282. \$27.50.

ONE of my vivid childhood memories is of my mother filling large jars with water and labelling them "October, 1962". She placed these, along with a large stock of canned goods, in a small room in our basement. We also assembled a box of canned foodstuffs for me at my kindergarten, where all the children practised going into the basement and putting our heads between our knees. There, a few miles from the White House, the Cuban missile crisis was no joke. Because it is easy to exaggerate either the goofiness or the high anxiety of such pathetic attempts at civil defence, one is tempted to read all of postwar history as dominated dramatically by the nervous shadow of nuclear weapons.

There are different ways to assess just how nuclear fission has affected us all. Joop van der Pligt has written a social-psychological monograph about public reactions to civilian nuclear energy. Allan Winkler presents a sweeping history of civilian and military nuclear energy in the United States. Carole Gallagher has compiled a moving book of photographs and recollections of dozens of people affected by the US government's atomic bomb tests during the 1950s. Singly and together, the three books provide an appalling picture of government arrogance.

Gallagher's book, *American Ground Zero*, leaves the reader in baffled outrage. More than 400 beautiful pages introduce us to diverse men and women affected by fallout from the US government's testing of nuclear weapons in the deserts of Nevada: employees at the test site, Mormons trained to obey authority, foot soldiers ordered to charge toward mushroom clouds, ordinary citizens who lived downwind of the tests, families whose members never seem to die of anything but cancer.

The book's quotations are an indictment of various forms of authority. Mormon leaders taught that the US constitution was divinely inspired, so many of their followers saw their radiation-related illnesses as God's punishment. A military chaplain reassured soldiers, many of whom returned from their post-explosion manoeuvres bleeding from the eyes and other orifices, that "there is no need to be worried, as the Army has taken all of the necessary precautions to see that we're perfectly safe here . . . the fireball as it

ascends up into the heavens . . . is a wonderful sight to behold." Scientific experts also share some of the blame, although Gallagher also sought out those who began to criticize government policies. But, of course, in the end the overwhelming blame rests with a succession of federal and state agencies that consist-

scientists who pulled off astounding scientific and technological feats or the arrogance of government decision-makers who stubbornly ignored risks and discounted the health of the public. Surprisingly, Gallagher's photojournalism has a richer texture than this book by a professional historian. Perhaps it was inevitable that a century of history in only 210 text pages should be superficial. But it need not have been bland, or marred by small and unsettling factual errors (for example, several people's occupations are misrepresented; James Conant is called the president of MIT, not Harvard).

Indeed, a short book of this scope could have been justified if it had a sharp analytical point, which this book lacks, beyond occasional references to scientists' participation in public debates. When he



From *American Ground Zero* Dorothea Lange

Blind faith — Mormons (Church of the Latter-day Saints) at Sunday worship, southern Utah, 1953. Top secret documents designated these people, who lived in the towns and villages around the atomic test site, as "a low-use segment of the population".

ently misled the public about the dangers of radiation.

Gallagher fears that, as a nation, Americans have become unwilling to recognize the nuclear war that the US government has perpetrated against its own citizens. Her book will help us to see. But the effect is numbing. The bad guys are consistently bad, the victims modern-day Jobs. With relatively little political analysis or hope, we can almost believe, like the Mormons, that this is God's punishment.

Next to Gallagher's document of condemnation, Allan Winkler's history of American nuclear energy, *Life Under A Cloud*, seems almost Panglossian. His polished, measured prose does not fully capture either the heroism of the atomic

does try to summarize, Winkler sometimes even gets it wrong, for example, attributing the collapse of American civilian nuclear energy to the antinuclear movement and an increasing concern for safety, rather than to scandalous mismanagement and enormous cost increases.

Van der Pligt makes a serious effort to explain public attitudes, especially of those opposing it, towards civilian nuclear energy. In contrast to the Americans of Utah and Nevada whom Gallagher describes, large numbers of people (Americans and others) have regularly protested against military and civilian nuclear energy. Relying on social-psychological studies from around the world — *Nuclear Energy and the Public* is intended as a

summary of this research — van der Pligt insists on the rationality of those suspicious of nuclear reactors, which exemplify risks that are invisible, involuntary, uncontrolled, uncertain, delayed and irreversible — all traits that inspire dread. This technology, it seems, had characteristics especially likely to make people nervous.

Studies of risk perception, initiated in the 1970s partly to understand why members of the public could be so irrational as to reject nuclear energy, have contributed a great deal to our understanding of public opinion and attitude formation. But van der Pligt's discerning if plodding use of this literature retains a flaw from the cognitive psychology that long dominated risk studies: a blindness to the formal organizations and political contexts shaping attitudes. Other than the frequent "credibility problems" of government agencies — meaning that they often lie to the public — van der Pligt ignores the ways that democratic procedures matter to people, that nuclear programmes come

to be associated with political parties and agendas, that attitudes are ways to take sides in public debates.

Van der Pligt's claim that the public is rational combines easily with the evidence provided by Gallagher and Winkler. That many citizens of the advanced industrial world distrust nuclear fission follows from the pugnacious way in which their governments have imposed it on them, an anxiety only reinforced by the technology's riskiness, itself long denied by those in power. Although many scientists have played important roles as independent critics since the 1940s, more have allowed themselves to become tools of their governments' enthusiastic promotional activities. The public might be willing to take more risks if they thought they could believe what the experts and officials were telling them about those risks. But so far, they have little reason to do so. □

James M. Jasper is in the Department of Sociology, New York University, New York, New York 10003, USA.

mechanics as our extragalactic visitor would know about condensed matter after a few days observing physicists cobbling together a grant proposal.

In fact, and Biagioli makes no bones about this, Galileo became a courtier (in the sense of having a court appointment) only when he was named "Philosopher and Chief Mathematician of the Grand Duke of Tuscany" in 1610. Galileo was then 46 years old and his best scientific work was behind him. Although his two main books, the *Dialogue on the Two Chief World Systems: Ptolemaic and Copernican* and the *Two New Sciences* were not published until 1632 and 1638, respectively, they were the fruits of his professorship at the University of Padua between 1592 and 1610. Galileo left the Venetian republic not to become a courtier in Florence but to have more time to arrange his material for publication. He wrote to the secretary of state of Tuscany on 7 May 1610: "If I could be repatriated, I would wish the Grand Duke to give me the leisure and the opportunity to draw my work to a conclusion without having to teach". Galileo got what he wanted: a professorship at the University of Pisa without obligation to give lectures or even reside in the city.

Biagioli recognizes that Galileo planned on enjoying the amenities of a research chair but argues that circumstances compelled him to change his role. The reader is invited to witness Galileo's "self-fashioning" into a courtier and to realize

From scientist to courtier

William R. Shea

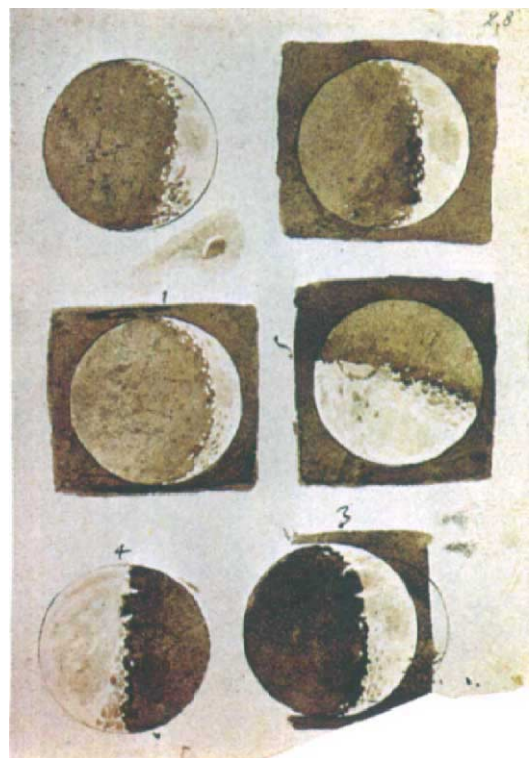
Galileo, Courtier: The Practice of Science in the Culture of Absolutism. By Mario Biagioli. University of Chicago Press: 1993. Pp. 402. £23.95, \$29.95.

A VISITOR from a distant galaxy arriving on Earth as scientists are busy filling in applications for research grants would form a peculiar idea of this delicate operation. Most scientists go through the exercise under duress, some in sullen silence, others in a stream of expletives. He would hear much about the waste of precious time and the arbitrariness, if not lunacy, of questions about the goal of their research, the equipment, methodology and technical support needed, and, last but not least, clearance from their ethics committee. He would discover scientists vetting friends' proposals to see how they will go down with the adjudicatory committee, and rumours might reach him of scientists offering to say a word in high places. He could not fail to notice that the pecking order in some disciplines is determined by the size of one's grant. On return to his galaxy, we would not be surprised if the visitor wrote an essay on terrestrial science in which grantsmanship figured prominently.

What an extraterrestrial had to say about our system of supporting research would be of the highest interest. Might he say of contemporary scientists what Biagioli says of Galileo? "The place of gifts within the logic of patronage explains the role of spectacular scientific produc-

tion in Galileo's career. Galileo needed to produce or discover things that could be used as gifts for his patrons." Are twentieth-century scientists tempted to do work that yields tangible results, however secondary, instead of tackling fundamental problems whose solution may bring no immediate relief to man's estate and so fail to qualify for the largesse of funding authorities? We cannot discount the pressures of contemporary patronage. On occasion, the research they want becomes the research we want simply because there is no other way of carrying on. The marvel is that so many scientists are willing to pass up the big bucks in the hope of a big breakthrough.

Every society has its own system of benevolent or despotic patronage. It needed someone as gifted and knowledgeable as Mario Biagioli to explore the relationships between science and government in seventeenth-century Italy. No historians of the period will leave this book without having learned something about the Tuscan court or the Roman *famiglia* (as the popes called their own entourage). But will they be greatly enlightened about Galilean science and the birth of the scientific revolution? My guess is that they will have learned as much about Galileo's



Biblioteca Nazionale, Florence; Scala

Galileo's sketches of the Moon pockmarked with mountains and craters, which were considered heretical for showing an imperfect heaven.

that "usually Galileo was not attacked because he was a Copernican but because of his (and his discoveries') extreme visibility and his success in becoming the mathematician and philosopher of the Grand Duke". In 1611 Galileo was welcomed at the Collegio Romano, the Jesuit University in Rome, where his telescopic observations were publicly praised — praise worth having, because the Jesuits, among whom were some of the best scientists in Europe, had made their own telescopes and confirmed Galileo's discoveries of craters on the Moon and satellites around Jupiter. We are now told, however, that Galileo's Roman Triumph "should be understood in terms of his connection to the Medici" and that he went to Rome as their "official envoy". The reason is that the government of Tuscany paid his travel expenses. On these grounds, shouldn't British scientists' lecture tours and attendance at meetings be "understood" in terms of the policy of the Royal Society?

Biagioli is more convincing when, in the footsteps of R. S. Westfall, he traces the relationship of Galileo with Prince Cesi, the founder of the Accademia dei Lincei, which Galileo joined in 1611, the year of his great success in Rome. As Galileo's patron, Cesi steered the *Letters on the Sunspots* (1613) and *The Assayer* (1623) through the shoals of Roman censorship and had them published at his own expense. Galileo hoped that he would also smooth, financially and otherwise, the path of the *Dialogue on the Two Chief World Systems*, but the Prince's death in 1630 deprived him of this influential friend and protector.

Biagioli's emphasis on patronage is also helpful in understanding the reaction of Pope Urban VIII to the publication of the *Dialogue* in 1632. The pope had treated Galileo with warmth and friendliness, and he expected to be repaid with courtesy and deference. The Roman censor had enjoined Galileo to mention the pope's view on the relativity of scientific theories and the impossibility of knowing with absolute certainty whether a given theory is more than a useful hypothesis. Galileo complied to the extent of stating this opinion (which he ascribes to "a most learned and eminent person") at the end of his book. Unfortunately, he puts it in the mouth of Simplicio, the Aristotelian pedant who cuts such a sorry figure throughout the entire *Dialogue*. Pope Urban VIII was personally affronted and bore Galileo a grudge to his dying day. Such bitterness might seem unbecoming for the Head of Christendom; Biagioli shows that it was comprehensible in a seventeenth-century patron. □

William R. Shea is in the Philosophy Department, McGill University, 3690 Peel Street, Montreal, Quebec, Canada H3A 1W9.

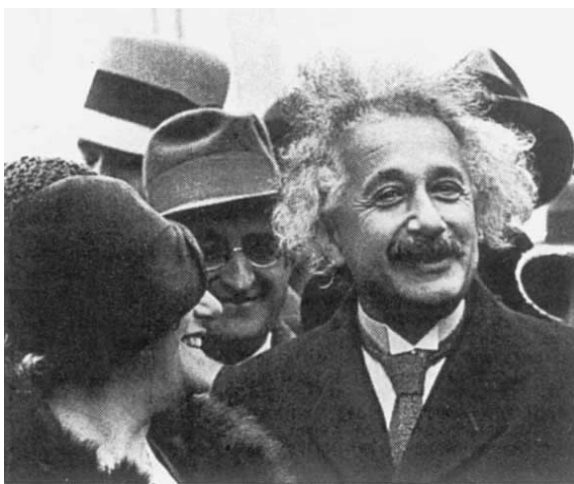
Relative values

Arthur I. Miller

The Private Lives of Albert Einstein. By Roger Highfield and Paul Carter. *Faber and Faber*: 1993. Pp. 355. £14.95.

LIFE can be complicated, especially in matters of the heart. So we ought not to be surprised that this was also true for Albert Einstein. Roger Highfield and Paul Carter have used Einstein's personal correspondence, some of it discovered as recently as 1986, to produce a narrative that reveals him to have experienced *angst* and joys not uncommon to most mortals.

But would this book have the same



Einstein with his second wife and distant cousin Elsa.

interest if we removed the name Albert Einstein? Let's do that and see what remains. We have the story of a pre-college student who falls madly in love, writes syrupy love letters to his beloved and courts her vigorously until she professes her love for him. Having achieved his desired end he beats a hasty retreat, preferring action at a distance. Finally the going gets too much and he ends the relationship.

Our 'hero' goes on to university where he falls in love (again) and a tumultuous love affair ensues, love letters and all. Being no Lord Byron, these letters are essentially the same as the previous ones. So is his preference for action at a distance once his girlfriend falls madly in love with him. He has doubts but doesn't want to lose her, preferring passionate reunions to a day-to-day relationship.

Both parents strongly oppose the liaison, his more than hers. He graduates from university, she does not. The plot thickens when the girlfriend becomes pregnant and, though he seems mildly enthusiastic about the baby, work comes first. They put the child up for adoption before he does the 'decent' thing and

marries her. Their close relationship is renewed with her totally dependent on him but he buries himself even deeper in his work. Then another child arrives, she becomes jealous of his friends and of any extra-familial activities and now the couple who could not live apart discover they cannot live together. He becomes unexpectedly successful and is offered positions elsewhere and expects her to move with him along with their two children. He feels suffocated by the relationship, while she takes refuge in the children. Sounds familiar? Let's continue.

A distant cousin 'on the prowl' arranges a meeting and the married man indulges himself. A liaison follows, as do love letters of course. A major position opens in the cousin's home town, which he accepts; the wife refuses to go. After the divorce our by now fantastically successful figure marries the cousin. He becomes bored and is distracted by the attentions of other women. The middle-aged man embarks on affairs — some with love letters, of course. Meanwhile the children from the first marriage have nothing but scorn for the second wife, seeing in her the miseries of their mother, still in love with the image of the passionate young man from her vanished youth.

With Albert Einstein as the central figure, this scenario takes on added interest. Einstein did more than create scientific theories that changed our view of nature; with Sigmund Freud he created the twentieth century. What is so amazing is that Einstein's most creative years were the ones spent in the Swiss Federal Patent Office in Berne, 1902–09, where he worked full time in addition to publishing about 50 scientific papers. Among them are the special theory of relativity and the groundwork for the general theory. Consequently, our eyes widen when, in the midst of a love letter to Mileva Marić in 1901, his college sweetheart and then first wife, he mentions "our work on relative motion". Passages such as this have led to the suggestion that Mileva deserves some of the accolades for special relativity. The authors review conclusive evidence to the contrary.

Highfield and Carter have an agenda: to smash an icon. So, for example, they spend an entire chapter on the 16-year-old Einstein's relationship with his first girlfriend, accusing him of being hypocritical for rejecting her pleas to spend more time together and then citing hearsay evidence from years later that he bears the blame for her subsequently unhappy life.

Bildarchiv Preussischer Kulturbesitz

Similarly, Einstein is linked indirectly with Mileva's academic failures. There is even a suggestion by the authors that Einstein beat Mileva. Sometimes the authors cite conflicting evidence for Einstein's extramarital liaisons during his second marriage to cousin Elsa. For example, after recounting a number of Einstein's liaisons during the 1920s in Berlin they quote Konrad Wachsmann, the architect who designed Einstein's summer house at Caputh (where some of the romantic meetings occurred), to the effect that the liaisons were "almost without exception" platonic. The authors do cite convincing evidence for certain of Einstein's affairs. In general their strategy is to weave selected passages from correspondence with reminiscences of Einstein from sometimes more than 40 years earlier in order to paint an unflattering picture.

Let's assume all this is true. Does it add to our understanding of Einstein's creativity or his science? This point is never addressed. Highfield and Carter are uninformative on such key issues and generally on what Einstein was up against scientifically in 1905. As a consequence, their narrative lacks the key line that they are telling us about a central figure in the history of ideas. It is like writing about Winston Churchill's life during 1940–45 without mentioning the Second World War. The authors offer us a disembodied figure, who becomes the subject of tabloid sensationalism. After all, maybe the worst possible scenario for Einstein's personal life was essential for his creativity, like Picasso's?

The titillating sensationalistic aspects of Einstein's correspondence aside, the critical portion for serious researchers interested in scientific creativity and not gossip, innuendo and rumour, are the early letters to Mileva (snippets are quoted by Highfield and Carter along with their interpretations — for the complete letters see J. Renn and R. Schulmann (eds) *Albert Einstein and Mileva Marić: The Love Letters*, Princeton University Press, 1992; for a review, see *Nature* **360**, 377–378 (1992)), in addition to other material relating to his life during 1902–09. With proper analysis they could reveal how his personal life affected his scientific work. Is this not what we expect of the artist, musician or writer? Einstein's life as a young man is the stuff of which movies are made. □

Arthur I. Miller is in the Department of History, Philosophy and Communication of Science, University College London, Gower Street, London WC1 6BT, UK.

■ *Einstein: A Life in Science* by Michael White and John Gribbin has just been published by Simon and Schuster, price £16.99. □

Science's new language

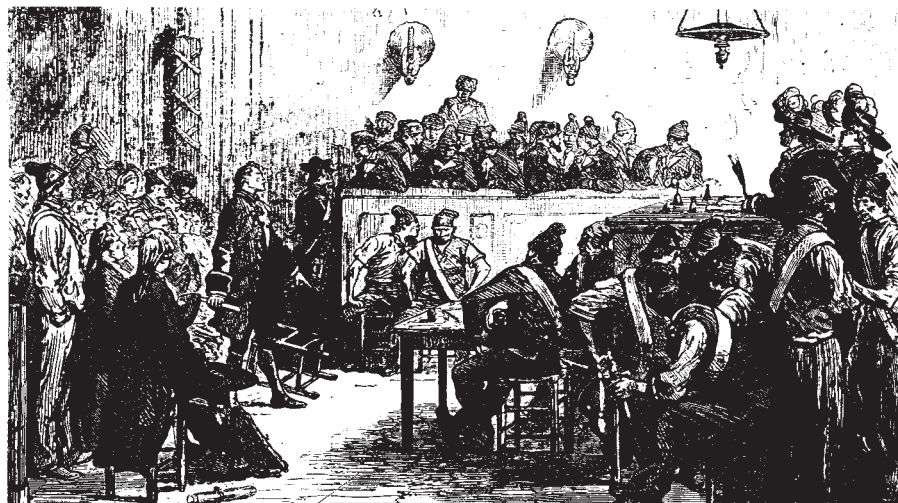
David Knight

The Enlightenment of Matter: The Definition of Chemistry from Agricola to Lavoisier. By Marco Beretta. *Watson Publishing International*: 1993. Pp. 396. \$49.95.

LAVOISIER was beheaded in May 1794, and the event will be marked, if not celebrated, by conferences and publications, of which Marco Beretta's is one of the first. He is an Italian working in Sweden and writing in English, and has an excellent command of the French sources; he is therefore excellently placed to emphasize the new language that Lavoisier provided for chemistry, and its spread throughout Europe in the years

stresses the differences in the terms and very language used before and after some heroic figure has transformed a science, accounting for earlier anomalies and persuading his colleagues to see things his or her way. This makes Lavoisier's revolution, where a new language was devised to transform a Baconian science into one having a clear and deductive structure, particularly instructive — perhaps indeed paradigmatic. Certainly Lavoisier's approach, example and language formed what Kuhn calls the paradigm within which the next generation worked.

Nevertheless, most students of the period have concentrated on Lavoisier's demonstration that the older view, that anything that would burn contained phlogiston, was incoherent. He replaced this with the theory that burning meant combination with oxygen, which had at first been called vital air, or eminently respirable air, or dephlogisticated air. Beretta emphasizes Lavoisier's respect



Lavoisier before the revolutionary tribunal that sentenced him to the guillotine in 1794.

between 1787 and 1800. Torbern Bergman and Guyton de Morveau have both in the past been given much of the credit for the new nomenclature, and Lavoisier's associates preferred to think of the new chemistry as a collective French achievement. But Beretta sees the achievement as essentially Lavoisier's, and the new language as crucial rather than as a convenient appendage to a new theory.

Lavoisier's great *Traité élémentaire de chimie* appeared in 1789, and he was very conscious that his chemical revolution was simultaneous with the political revolution around him. This eventually cost him his life, as a prominent and extremely wealthy tax collector. Indeed, this was the first major transformation in the sciences to be described by participants as a revolution. The idea that such sharp discontinuities, like the catastrophes prominent in the geology of Lavoisier's period, make up the history of science is a crucial feature of Thomas Kuhn's philosophy of science. He

for G. E. Stahl and his phlogiston theory, which had brought order into chemistry and formed a necessary stage in its development; and his preoccupation with scientific language. In the *Traité*, there is reference to E. B. Condillac and his writings on language, which were developed from those of John Locke; but for Beretta this is a clue to a most important part of Lavoisier's thinking.

In 1787 the treatise *Nomenclature chimique* was published, with Guyton's name first in the list of authors, and Lavoisier's second; it has often been supposed that this indicated their importance, but Beretta shows that the four authors were in fact listed in order of age. For him, this book represents Lavoisier's scheme, beautifully calculated to make the new theory palatable, and indeed almost inevitable; but very bold. One of his exemplars was Linnaeus, whose system of naming plants and animals had brought order into natural history. Linnaeus based his procedure on external characters:

Mary Evans

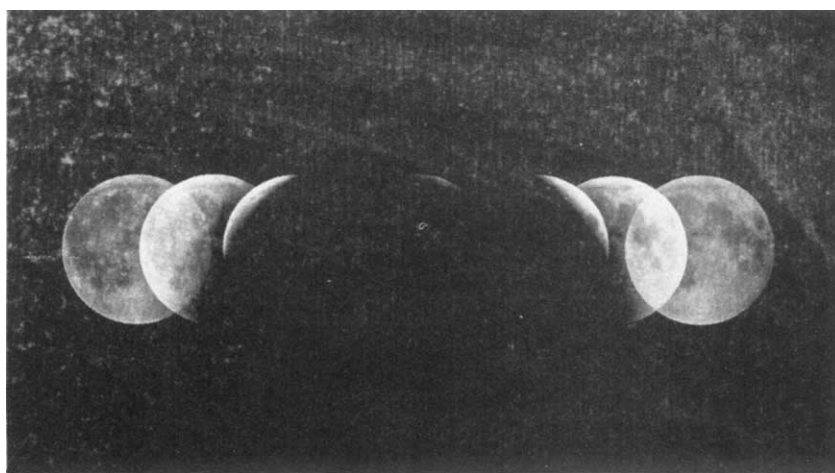
Lavoisier wanted to go deeper.

Here indeed he went astray famously over oxygen, which he named from Greek roots believing it to be the generator of acidity. His etymology came in for some attacks, but much more important was the basis for this name in a generalization from the known acids. First hydrogen sulphide and hydrogen cyanide, and then much more seriously the very strong acid from sea salt, turned out to contain no detectable oxygen when carefully analysed in the absence of water; Humphry Davy in 1810 concluded that they contained none. Oxygen was therefore a misnomer, and the French theory "the baseless fabric of a vision": to accuse the French of calling up cloud-capped towers and gorgeous palaces when they should have been doing sober science was a great pleasure to an Englishman and his listeners in wartime.

All science is open to falsification, and even Lavoisier did not get it all right; but Beretta is critical of Britons in the eighteenth century who saw language in science as a matter of conventions only, and chemistry as essentially a mass of facts. He is very interesting on the different national traditions in Europe in Lavoisier's day, and the way these affected the reception of the new chemistry. In Britain, the language was played down at first, but by 1800 its sheer convenience had prevailed. But perhaps we should note, as Beretta does not, that in medical practice 'antiphlogistic' remedies for reducing fever persisted into the 1820s, and indeed Davy was given them in his last illness. Priestley, though politically pro-French and, therefore, ending his days as an exile in the United States in 1804, was firmly in favour of the phlogiston theory and nomenclature; but his example does not seem to have kept most English speakers from going along with the French. In Germany, political considerations were more important, and the French armies brought revolutionary principles in politics and chemistry with them.

The book ends with an appendix on alchemical imagery, with the message that this had nothing to do with chemistry. Certainly Beretta's ideal, like Lavoisier's, is a clear language, where analogies rather than metaphors are brought out. But we may feel that something was lost when the language of chemistry became so emphatically prosaic and algebraic, in accordance with Condillac's ideal; Samuel Taylor Coleridge went to Davy's lectures to improve his stock of metaphors, but the language was losing its resonance already. Beretta's book touches on many features of chemistry before 1800, and is a stimulating addition to the literature. □

David Knight is in the Department of Philosophy, University of Durham, Durham DH1 3HN, UK.



MULTIPLE exposures of the Moon with the camera tracked at the rate the stars move across the sky reveal the shape of the Earth's shadow. This is one of many examples in a clear and practical guide to eclipse photography, *The Cambridge Eclipse Photography Guide: How and Where to Photograph Solar and Lunar Eclipses*. Cambridge University Press, £10.95, \$16.95.

Discovery of the century?

Joseph Silk

Wrinkles in Time. By George Smoot and Keay Davidson. Little, Brown: 1993. Pp. 321. £18.99.

Ripples in the Cosmos. By Michael Rowan-Robinson. Freeman: 1993. Pp. 224. £16.99, \$22.45.

Afterglow of Creation. By Marcus Chown. Arrow: 1993. Pp. 171. £5.99.

RARELY in the modern history of science has there been such a stir as that caused by the Cosmic Background Explorer (COBE) satellite discovery of ripples in the cosmic microwave background radiation. Front-page headlines in newspapers around the world heralded this announcement, accompanied by a picture that, it only later transpired, was mostly instrumental noise. Within hours of seeing the headlines on his way to Narita Airport in Tokyo, John Brockman, the high-profile literary agent, was on a payphone to George Smoot, principal investigator of the Differential Microwave Radiometer (DMR) experiment that mapped the ripples. A book was duly conceived, aided by "the largest deal in the history of science publishing" according to Marcus Chown, and has promptly been written, in collaboration with science journalist Keay Davidson. *Wrinkles in Time* is the outcome, a somewhat personal series of anecdotes about the vagaries of space astronomy.

What is remarkable about the DMR is that with off-the-shelf (in 1975) technology, Smoot was able to make two major breakthroughs in cosmology. He played a leading role in the group that flew the DMR aboard the U-2 spy plane, to spy

on the heavens and discover the motion of our Galaxy relative to the cosmic microwave background radiation. The measured dipole anisotropy provided the first proof that the microwave background was of truly distant origin and likely to be the relic fireball from the Big Bang.

The slight deviation due to our motion was 1 part in 1,000. Theorists had predicted that at a level of 1 part in 100,000, one ought to be able to detect the blemishes in the microwave background that represent the seeds of large-scale structure. One sees the primordial fireball radiation when the Universe was only 300,000 years old; it provides a glimpse of the Universe before structure had formed. However, this increase in sensitivity required a space-borne platform, and Smoot nurtured the DMR experiment into a reincarnation on board the COBE satellite. The satellite was originally destined for the Space Shuttle, but the Challenger disaster made necessary a redesign and the experiment had to be reduced to fit into a Delta rocket that was launched four years ago. Fifteen years had elapsed since NASA's decision to proceed with COBE.

The wait was worthwhile. The far-infrared absolute spectrometer (FIRAS) carried on COBE, designed by John Mather, measured the most precise fit to a Planck function ever performed on a black body in the sky. The cosmic microwave background radiation is so perfect a black body that it could have originated only in the fiery furnace of the first months of the Big Bang. The DMR also performed according to plan. It mapped the entire sky at a resolution of 7 degrees, and measured slight deviations from uniform-

ity in the microwave background. The discovery of fluctuations, at the expected level of 1 part in 100,000, meant that the Big Bang was finally a respectable and consistent theory: the last remaining enigma was solved.

The vagaries of space astronomy are also the subject of *Ripples in the Cosmos* by Michael Rowan-Robinson. The author is professor of astrophysics at Imperial College, London, and has written several popular books on astronomy. He played an important role in the Infrared Astronomical Satellite (IRAS), an infrared space telescope built by a team from the United States, United Kingdom and the Netherlands and launched in 1983. IRAS provided the first all-sky map at infrared wavelengths, and revolutionized our perspective on the cosmos. The most luminous galaxy in the Universe was first seen as an IRAS source. Only in the infrared can one penetrate the cores of interstellar dark clouds, where stars are conceived and born. Only in the infrared can one get an accurate all-sky tally of the large-scale galaxy distribution. This map was used to provide a measure of where in the Universe the gravitational acceleration originates that accounts for the motion of our Galaxy as reflected in the dipole anisotropy in the microwave background. Rowan-Robinson includes a chapter that describes the impact on his life of the ripple announcement.

For the real dirt on COBE, however, I recommend Marcus Chown's slender tome, *Afterglow of Creation*. The science editor of *New Scientist* brings a journalistic fervour to delve into the personalities behind the headlines. Here you will read random quotations by the key players about the many rivalries between scientists striving for their share of glory. Both Chown and Rowan-Robinson take care to give proper credit to Arno Penzias and Robert Wilson for accidentally discovering, while employed by Bell Laboratories as young radioastronomers, the microwave background radiation. Chown describes the discovery of the microwave background and the COBE announcement of the ripples, peppered with speculations about Nobel prizes, and why scientists, when stumbling for appropriate words, inevitably need to invoke God for an appropriate metaphor. The public was bombarded with deistic analogies in the post-COBE euphoria, including seeing "the face of God", the "handwriting of God", and discovering the "Holy Grail", not to mention the more secular tribute of "the discovery of the century, if not of all time". Most of the scientists who quoted these words to the media soon came to regret their moment of passion, as a backlash inevitably developed. However, now that the DMR ripples have been confirmed, there is little doubt that the discovery of the cos-

mic microwave background fluctuations, as well as the unprecedented precise measurement of its black body spectrum, are a major breakthrough in our understanding of the cosmos. We can reasonably anticipate that appropriate recognition will be made when the attention of the Nobel committee refocuses on astrophysics in a few years' time. □

Joseph Silk is in the Astronomy and Physics Departments and Center for Particle Astrophysics, University of California, Berkeley, California 94720, USA.

Physically Impossible

David Goodstein

Fear of Physics: A Guide for the Perplexed. By Lawrence M. Krauss. *Basic Books*: 1993. Pp. 206. \$20. (To be published in the UK by Jonathan Cape in 1994 at £17.99).

WHAT we have here is nothing less than an attempt to do the impossible. *Fear of Physics*, with an apology in the preface to Erica Jong, is subtitled *A Guide for the Perplexed*, Moses Maimonides unattributed. I'm sure Erica could have taught Maimonides a thing or two, but that's not what this book is about. This is definitely Maimonides, gently introducing Erica to the wisdom of the ages, meaning, of course, contemporary theoretical physics.

Lawrence Krauss, of Case Western Reserve University, has undertaken to explain it all in simple language, almost without committing a single equation. In less than 200 pages, he covers just about everything. Scales, dimensions, little numbers and big numbers and the usefulness of scientific notation; from Galileo to Gell-Mann, from dark matter to critical point fluctuations to the Higgs boson condensate; it's all here (I didn't notice anything on the fractional quantum Hall effect, but I may have dozed off for a few pages). Every bit is explained in clean, simple prose. There is barely a paragraph anywhere that could have been made easier to understand. And yet, sad to say, the enterprise as a whole doesn't quite succeed.

Part of the reason has to do with Krauss. His accounts of physics are technically excellent, but he tends to be sloppy about anything less important to him than physics itself. For example, in one place we are told that Johannes Kepler was Tycho Brahe's student. At a different level of intellectual sloppiness, we are told at another point that Albert Einstein was the Columbus of the twentieth century (you know, the world is round, and spacetime is

curved). Another problem is that Krauss feels obliged to intersperse amusing anecdotes in his text, but he has no flair at all for telling them. He chooses the oldest chestnuts around, then delivers them poorly. That includes the one about the theoretical physicist's contribution to dairy farming, the spherical cow. He immediately admits that it's not funny, although he makes it the central metaphor of the book (there's a picture of a cow on the dust jacket — let's hope Erica doesn't misunderstand). I've heard the joke many times and, depending on the skill of the storyteller, sometimes it is funny.

These are quibbles. The real problem is the impossibility of the task itself. Krauss explains point A to us; he can't explain it to us the way he would to an undergraduate class, much less to his own colleagues, because that is not the nature of the undertaking. Instead he surrounds it with metaphors, analogies, simplified examples and so on. When he's done, because he does it very well, we have a pretty good idea of why physicists think point A is so important, and we may even think we've grasped the point itself.

The problem comes when he starts to explain point B, which requires that we already understand point A. He uses all the same strategies just as nicely as before, but now we realize that point A was more slippery than we thought, and point B is really shaky. Then it's on to point C. By the time we get through the entire alphabet pretty much all is lost, except to the perplexed who had no fear of physics in the first place.

By ironic coincidence, I read Krauss's explanation of the importance of the Superconducting Super Collider on the very day that it died ("Congress finally drove a wooden stake through its heart", said the *New York Times*). Some day, when history sorts out all these things, it may be decided that the era that ended that day failed for the same reason Krauss's book did. We physicists left the public 300 years behind. The gap has turned out to be too big to fill in.

Richard Feynman, who is one of Krauss's heroes, once had a similar failure. He tried to teach physics to a class of bright first-year college students. He ended up instead leaving us a set of books that are useless to students but are nevertheless enduring classics of the scientific literature. *Fear of Physics* is not in that category, but (aside from my quibbles) it is beautifully written and painless to read. Somewhere on the spectrum of scientific sophistication, there must be an audience for this book. I hope the book and the audience find one another. □

David Goodstein is in the Department of Physics, California Institute of Technology, Pasadena, California 91125, USA.

A natural education

Andrew Whiten

Feral Children and Clever Animals: Reflections on Human Nature. By Douglas Keith Candland. Oxford University Press: 1993. Pp. 411. \$30.

ON 17 October 1920 near Midnapore in India, the Reverend A. J. Singh sought to drive out the ghosts that local people reported to inhabit a wolf's den. The mother wolf was shot and the den was dug out. Singh's diary recorded what they found: "the two cubs and the other two hideous beings were there in one corner, all four clutching together in a monkey-ball. The ghosts were more ferocious than the cubs, making faces, showing teeth, making for us when too much disturbed, and running back to form the monkey-ball." By this account the girls, later named Kamala and Amala, were true wolf-children. Together with the other 'wild' or feral children who occupy the first part of Candland's book, these wolf children are inevitable objects of curiosity for all those who seek to delineate the bounds of human nature.

Curiously, the mirror-image of such nightmarish scenarios is with us daily as we rear puppies in our homes, without undue concern about the implications for understanding either dog or human nature. Yet, when the species to be human-reared is an ape with a brain more like our own, excitement grows again about the prospects for understanding the scope of our own nature in the comparison. Attempts to home-rear apes and teach them language form the fourth and final part of the book.

Between these in-depth accounts of feral children and civilized apes, Candland sandwiches analyses of a gamut of other studies from the past century, which offer to illuminate human nature, typically through human-animal linkages and comparisons. Through the surprising diversity of studies that Candland binds to this central task, he succeeds in another major achievement — an eye-opening history of psychology. Candland's meticulous scholarship repeatedly reveals subplots and obscure yet important details surrounding famous studies that have all too often become distorted or diminished in contemporary textbooks. Thus, for example, when we hear of Pfungst's well-

known dismissal of counting by the horse 'Clever Hans' as merely cued by his trainer von Osten, we also learn that Pfungst's hypothesis was never actually tested at the time; that after the famous critical work by Pfungst, Hans joined other horses in a series of more complex experiments conducted elsewhere by Karl Krall; that Rendich made comparative studies of count-



Spelling it out—Richard L. Garner teaches a chimpanzee to turn alphabet blocks. One-hundred years ago, he made the first observations of chimpanzee and gorilla behaviour. He claimed success in teaching an orphaned chimpanzee to make limited human-like speech sounds; he also invented the technique of recording natural primate vocalizations and playing them back to the animals to discover their meaning.

ing by Hans and a dog; and that further efforts to teach dogs to speak or communicate using signs anticipated contemporary efforts to teach apes sign language by over half a century.

But Candland has another major theme: that whatever the conclusions of the studies discussed, the particular ways in which investigators frame their questions tell us a great deal about what it is to be human. The thinking of human beings who act as investigators shows them to be children of their age and culture: what pre-darwinians expected to learn from feral children was fundamentally different from later preoccupations with their 'animal instincts', for example. But complementing this cultural relativity is at least one constant theme that addresses what it is to be human: the drive to worm one's way into others' minds. A pre-occupation that surfaces repeatedly is how to reveal the psychological experiences of those who are congenitally dumb —

whether clever horse, ape, feral child or a child such as Laura Bridgman, totally blind and deaf. In the case of a feral child such as the 'wild boy of Aveyron', the idea of investigators was that if he could only be taught to speak, he would be able to say what it had been like to be a human being living in a 'state of nature'. In modern times the same heady objective has been espoused for what we may learn of the mental world of an ape that is taught language. And in between, Candland helps us to see the same goal being nursed in investigations ranging from that of Clever Hans to that of Little Hans, for in the latter, Freud equally endeavoured to establish communication with a normally silent mind, the elusive subconscious which in the case of the boy Hans he saw as identifying the father with an animal, the horse of which the boy was terrified. This psychoanalytic reference to the animal-in-man is yet another illustration of the complex web of themes that make this a highly original and thought-provoking book.

From *Feral Children and Clever Animals*

I suspect however, that many readers will, like myself, find too much undigested detail, coupled with a tendency to avoid setting out the specific conclusions we should be left with after such a rich and varied meal. This may be partly deliberate: one feels that Candland's strategy is to direct our attention to neglected issues, make wise comment, then sit back to let the reader struggle to his or her own conclusions.

Two sets of studies that are not referred to could have deepened our understanding of what light is cast on the nature of a species when its members are reared in isolation or by other species. One is work on the ontogeny of bird song, which, through sophisticated variations on 'Kaspar Hauser' deprivations and cross-fostering, has greatly furthered our understanding of nature-nurture interaction. The other is Mason's rearing of monkeys with dog foster-mothers — the closest to a controlled and systematic version of the 'wolf-child' as we are likely to see. The monkeys acquired no obvious dog behaviour (which appears to contrast markedly with the wolf-like behaviour of Amala and Kamala), yet they lacked significant parts of what we should otherwise perceive as monkey nature. □

Andrew Whiten is in the School of Psychology, University of St Andrews, Fife KY16 9SU, UK.

Sound stuff

David Pye

The Sonar of Dolphins. By Whitlow W. L. Au. Springer-Verlag: 1993. Pp. 277. DM 148, £58.50.

How must the world appear to other species, whose senses differ in varying degrees from our own? It is hard to imagine what it must be like to use natural echolocation, at which dolphins and bats excel. Although we now have a thorough understanding of echo principles and can

comprehensive review of the dolphin story, and a 'batty' equivalent is devoutly to be desired.

For Whitlow Au has written a splendid book, which is likely to become a classic in its field, and of considerable interest well outside it. This is not only an authoritative reference work but it also contains a clear historical perspective and a treatment of all the fundamentals for a reader new to the subject. There is no excuse afterwards for thinking a decibel is a unit of sound intensity or for confusing peak-to-peak with root mean square amplitudes. (But is it merely an unfortunate slip to say, contradicting the adjacent equation,

that sound pressure level, rather than intensity, decreases with square range?) The succinct accounts of time versus frequency domains, Fourier transformation, near and far fields, psychophysical methods and detection theory should all do much good.

There follows a detailed and systematic survey of dolphin ears, hearing and acoustic discrimination. Sound production, emission and directionality lead into analysis of the signals themselves and a comparative account of the pulses of 21 species (although echolocation has not actually been demonstrated in all of these) out of a possible 65. Inevitably parts of this will be seen by some as contentious, for matters lacking firm evidence have often given rise to heated dispute, some of which cannot yet be resolved — even the source of sound within the dolphin's head is still a mystery. But the author deals fairly and confidently with the confusion and conflict, presenting evidence and opinion in an admirably even-handed way that is both stimulating and reassuring.

Eventually the stage is set for reviewing the very impressive capability of dolphin echolocation, followed by its interpretation by acoustic theory and speculations on the neural processing involved. Again, some of this may be contentious but that is not the point — its stimulus will be justified if it leads to more than verbal dispute. At least this is a new starting point.

Finally Au takes a careful look at bat systems for comparison and concludes with a brief, moderate consideration of outstanding problems, deficiencies of past approaches and indications of future potential. The whole is beautifully produced and copiously illustrated with admirable clarity. □

David Pye is at the School of Biological Sciences, Queen Mary and Westfield College, London E1 4NS, UK.

Thinking in circles

Susan Greenfield

The Creative Loop: How the Brain Makes a Mind. By Erich Harth. Addison-Wesley: 1993. Pp. 196. \$21.

SCIENTISTS are finally waking up to the question of consciousness. The nature of the mind, for so long regarded as the terrain of philosophers, psychiatrists and cognitive psychologists, has become a growth industry among those with a detailed knowledge of physics and biology. Like most scientists who have produced books on the subject, Harth is not a neuroscientist; he is a physicist who nonetheless succeeds in keeping his background well under control. The chapters follow each other effortlessly in a witty yet authoritative style that should immediately capture the imagination of the general reader.

Harth starts with a refreshing departure from other forays into the mind-brain problem by reflecting on the frequently neglected question of the actual nature of life itself, and hence of the divergent development of animate, purposeful organisms away from inanimate objects. This approach brings a new perspective on the hoary issue of the physical versus the mental: the author focuses attention on the difference in objective time and space, where there is only local causality and isolated entities in a perpetual present, in contrast to the vast spans in time and space by which we bond together a subjective universe. In attempting to relate the objective to the subjective, Harth espouses a 'physicalist' standpoint: this approach, which has already been broached by such diverse authorities as the mathematician Roger Penrose and the philosopher Thomas Nagel, is that science may ultimately account for the talent of our brains to translate the physical into the phenomenological. But the current laws and relations that might be thought adequate by the 'materialist' to explain the mind will need to be supplemented by as yet unknown principles and laws.

The central question around which this book is written is basically that already publicized by Dennett as the 'Cartesian Theatre', the fallacious anatomical locus of consciousness. Like most who have contemplated this issue, Harth rejects the idea of a hierarchical pyramid of neurons leading to a final set of 'gnostic neurons' as a centre of consciousness. By contrast, he is reluctant to embrace the only apparent alternative, parallel processing. He rejects the idea of transiently salient 'multiple drafts' or 'bundles' of consciousness as ducking the issue of the undeniable cohesion and continuity of our individual awareness. If we 'demystify the brain' by



Fathoming the mystery.

produce excellent sonars for use in water and radars for air and space, there are still many problems in understanding the sonar of animals.

Sonar study itself has never been cohesive. Quite apart from the obvious differences between the size and needs of dolphins and bats, there are fundamental differences in the way they operate. These are probably associated less with their separate evolution than with differences between air and water as acoustic media, which also demand different techniques in their study. So there have always been 'bat people' and 'dolphin people', who only meet at the splendid but infrequent joint conferences run as NATO Advanced Study Institutes. Even the literature on dolphins often seems (to a bat person) to be in quite obscure publications, so the conferences were occasions of delight as everyone discovered what the 'other side' had achieved. Now at last there is a

claiming it is like an anthill, then surely we have only 'mystified the anthill'.

Harth's solution is inspired by the observation that neurons are always relaying signals, never merely receiving them. The centrepiece of his theory is thus that reverberating loops of neurons are responsible for mental phenomena. In particular he makes much of the back projection from the visual cortex to the lateral geniculate nucleus (LGN) of the thalamus, an area that neuroscience neophytes normally view as a simple stepping stone in a one-way path up into the cortex. The obsession with this particular thalamo-cortico-thalamic loop inevitably makes Harth focus much of his contemplations on consciousness in terms of eidetic visual awareness, 'pictures in the head'. He accounts for the subjective quality of our vision as being at the behest of the signals bombarding the LGN from the 'higher' regions of the brain where memories, prejudices and the like will be generated. The more dominant this 'higher' input to the LGN, and hence the more unfettered from the sobering input of our retina, the more we will be at the mercy of our dreams, fantasies and delusions.

As it stands, this vision is not completely persuasive. Harth admits that the physicalist standpoint cannot yet account for the phenomenon of the feel of consciousness. Although this is a completely respectable and acceptable limitation, we might have expected, nonetheless, a richer account of the actual physical-chemical events in the brain itself, beyond the simple game of shuttlecock between relatively extensive regions of cortex and thalamus. Once he strays from this particular path, however, Harth becomes rather lost. He mentions the brainstem, yet does not elaborate on the way in which arousal might play a part in his scheme: he admits that beyond our vivid awareness of raw pictures, we would need to retreat into the 'neural jungle of the cortex': but then he leaves us there in uncharted territory, without any clue as to how more abstracted awareness might be subserved. He points to drugs and anaesthesia in support of the physicalist standpoint, yet he does not incorporate the actions of, say, LSD or morphine, let alone the powerful intricacies of neuromodulation, into what remains in consequence a factually thin, impoverished viewpoint.

Even when we look beyond the machinations of actual neurons, it would have been helpful for the description of awareness of 'pictures in the head' to be placed in the context of other candidates for the seat of consciousness, not least the highly popular circuit linking another region of thalamus (the nucleus reticularis) and the cortex, which has been previously proposed to provide the all-important ingredient for making us conscious of the

otherwise unconscious. Another disquieting whitewash is that Harth dismisses the degree to which our consciousness is kept in touch with external reality by glibly referring to 'an internal logician'. This is surely tantamount to stretching out in the front stalls of the Cartesian Theatre, and certainly offers no new insight. But then again, Harth makes no claim to be a neuroscientist and cannot therefore be expected to produce an embroidered account of the intricate workings of cortical circuits in literal, physiological terms. The onus is surely on neuroscientists to develop a scheme free of metaphor and extending beyond one or two pathways, be they linear or looped.

Freeman Dyson is quoted at the beginning of a chapter as claiming that, "In dealing with the problem of consciousness, physicists have had courage but no competence, biologists have had competence but no courage". The book frames questions that really only neuroscientists can answer: it is high time they proved their worth. In the meantime, Harth transmits an infectious concern for the 'fuzzy' edges that surround the beginning of life, a contempt for the potential of the plodding analogues of computation and a plea that we are not changing into automata bereft of original ideas. The strength of this book lies not in its detail, but in its breadth and its humanity. □

Susan Greenfield is at the Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK.

Spreading information

Jonathan M. Mann

The Slow Plague: A Geography of the AIDS Pandemic. By Peter Gould. *Blackwell: 1993. Pp. 228. £35 (hbk), £12.99 (pbk).*

PETER Gould takes the readers of his book on a voyage, proposing to offer a new look at AIDS, with new vistas and horizons but, unfortunately, he leaves us disappointed and dispirited. He starts with an attractive thesis — that geographical perspectives on the human immunodeficiency virus (HIV) pandemic may offer a new insight — but simply fails to deliver on his promise. This is particularly frustrating as he is addressing the critical question of how to give shape (and therefore meaning) to a complex phenomenon, the global epidemic of HIV and AIDS.

It is true that existing 'pictures' of the pandemic are generally simplistic, failing to capture the diversity, differential in-

tensity and velocity, and unstable nature of HIV/AIDS in the world today. As a relatively new historical occurrence, the HIV pandemic is still highly dynamic, which includes its continued spread in all affected areas (in the United States, 40,000–80,000 new HIV infections are projected this year, along with about 75,000 new infections in Europe), its spread to previously little-affected places (such as rural areas or South-East Asia) and its highly differentiated and evolving nature within single urban areas such as Miami, New York, Paris or Nairobi. For all these reasons, the promise of a new approach to structuring our understanding of the pandemic, based on an emphasis on the 'space' of the geographer, is appealing.

The book is indeed initially stimulating, with sharp and pungent writing. The author's wide-ranging observations and speculations are full of energy and passion. He shines when criticizing others, which, at least at the beginning of the book, heightens our expectations.

The first disappointment is Gould's curious weakness for minor inaccuracies and a lack of precision. For example, he is mistaken when he labels the US Centers for Disease Control (CDC) "the international center for reporting the outbreak of diseases all over the world"; states that Sweden proposes that most people with AIDS should be put on an island; and reports that "tens of millions of WHO and CDC research dollars" were spent in Zaïre.

Second, the reader becomes impatient wondering when geography's critical insight will finally be articulated. For example, the chapter on Thailand, although entertaining, could have been written by any number of authors uninformed beyond general literacy about the science of geography. There are some excellent and interesting maps of cumulative AIDS cases in the United States (particularly in the Bronx in New York) and brief discussions of spatially contagious and hierarchical diffusion, yet these do not add enough geography to satisfy the reader.

But let us come to the central point. Having promised and failed to demonstrate how geography will make a critical difference, the author's real agenda emerges. He informs us that there has apparently been a conspiracy against geographers; every time they have tried to clear up the confusion about AIDS, they have been rebuffed. The following is typical of the tone of his discussion: "I want to . . . look at a few . . . aspects of the way bureaucratic power, combined with a deadly combination of Establishment ignorance and arrogance, suppressed any consideration of the spatial dimensions of the epidemic, denying both the scientific community and the general

public any knowledge, insight and real awareness of what was unfolding into a human tragedy." Really? Regrettably, in the end, this is just an angry book, with familiar targets for the general reader: bureaucrats, governments and doctors ("... the medical profession, which itself seems to have been elevated to the sacred"), as well as scapegoats for AIDS "insiders": epidemiologists in general, the CDC and the National Institutes of Health.

Then at last, near the end, the author informs us about his central concern, which is that in HIV/AIDS prevention and care, the rights of certain individuals are taking precedence over the rights of the larger group. Finally, we are on familiar territory, even if geographical "new found land" is not involved.

The world does face a global crisis in the continuing spread of HIV and its rapidly mounting personal and social burden. How we define a problem largely determines what we do about it. Therefore, despite this book, I am convinced that creative efforts to 'map' this pandemic, including its social geography, could well be important in the evolving effort to give shape and meaning to the global challenge of HIV/AIDS. □

Jonathan M. Mann is Francois-Xavier Bagoud Professor of Health and Human Rights, Professor of Epidemiology and International Health, and Director, International AIDS Center, Harvard School of Public Health, Cambridge, Massachusetts 02138, USA.

Sex, lives and hormones

Peter Marler

Behavioral Endocrinology. Edited by J. B. Becker, S. M. Breedlove and D. Crews. MIT Press: 1992. Pp. 574. \$34.95.

FRANK Ambrose Beach effectively launched the systematic investigation of behavioural endocrinology 40 years before his death with his pioneering 1948 monograph on hormones and behaviour. Two years later he published his famous diatribe, *The Snark was a Boojum*, berating the so-called 'comparative' psychologists of his day for their myopic focus on the white rat, and urging them, the boojums, to extend their mandate to the animal kingdom at large. The discipline had to wait almost half a century for an all-embracing undergraduate text to emerge, a gap now handsomely filled by *Behavioral Endocrinology*, whose editors are all inheritors of the Beach tradition. He would surely have celebrated the degree to which his call for snarks to join forces and endorse the phylogenetic approach has been taken to heart as they explore the way in which hormones exert their manifold effects on the behaviour of organisms.

The 16 chapters by 19 authors, all experts in their field, range from sexual behaviour and sex differences in cognitive functions in humans (C. S. Carter, E.

Hampson, D. Kimura) to the wandering behaviour of the tobacco hornworm and other aspects of invertebrate behavioural endocrinology that illustrate well the advantages, if you are a reductionist, of working with simpler nervous systems (J. W. Truman). Most of the authors are psychologists, and almost inevitably rodents get the most intensive coverage, but the phylogenetic breadth covered is nevertheless impressive. A masterly overview of the behaviour of reptiles and fish (D. Crews) serves to remind us that the causal links between sexual behaviour and gonadal hormones are by no means inviolate, and have been severed repeatedly in the course of evolution. His essay serves as a first step in placing behavioural endocrinology in the functional perspective we associate with modern developments in ecological and evolutionary thinking, although more cross-references to this literature would have been welcome (one searches the bibliography in vain for references that students might want to pursue to the seminal contributions of such biological behaviourists as William Hamilton, John Krebs, John Maynard Smith and E. O. Wilson).

The book is expressly designed as an introductory text that begins from first principles and assumes a minimal background knowledge in biology. It is well designed, has a useful glossary and a bibliography with 1,143 entries. The introductions occasionally fall into the somewhat chatty 'gee whiz' style that sometimes plagues introductory psychology texts (for example: "amazing as it may seem" (p. 12); "the mysterious and fascinating world of sexual behavior" (p. 69); "hormones are not love potions" (p. 69)), but they quickly shed these stigmata of adolescence to grapple with an impressive breadth of subject matter, from molecules to cognition.

Much of the book can serve well as an introduction for students in the neurosciences. Some chapters, such as that of Crews, focus especially on ideas and principles, and others on details of the physiological machinery, such as those on male (M. J. Baum) and female (Carter) sexual behaviour. Chapters by S. M. Breedlove, D. B. Kelly and E. Brenowitz on sexual differentiation of the brain and courtship behaviour deal with sexual dimorphisms, the role of hormones in their development, and the influence of sexual differences on the adult social behaviour of mammals, birds, amphibians and fish. The distinction between activation and organizational effects of hormones on behaviour, fundamental though not necessarily categorical, pervades the entire book. Parental behaviour is authoritatively reviewed by J. S. Rosenblatt, with many illuminating comparisons between mammals and birds, and salutary reminders that some aspects of parental



MEXICAN wolves showing affection — the male (left) is giving the female a ritual, painless bite. Taken from *Trail of the Wolf* by R. D. Lawrence. Key Porter/Verulam, £17.95, Can \$34.95.

behaviour are independent of current hormonal controls.

The prime emphasis on sex is nicely complemented by chapters on nonreproductive behaviours, including a masterly account by R. M. Sapolsky of the psychosomatic relationship, both normal and pathological, between behaviour and the stress response. Other chapters review the role of hormones in the control of aggression (E. P. Monaghan and S. E. Glickman), feeding and drinking (E. M. Stricker and J. G. Verbalis) responses to the physical environment (R. Silver) and behavioural rhythms in oestrous behaviour, and circadian and circannual cycles (L. P. Morin and J. Dark). The appropriateness may be questioned of some of the more speculative material for an undergraduate text. A wide range of human sex differences in motor functions and cognitive abilities known to vary with phases of the menstrual cycle are assumed to have a direct hormonal basis (Hampson and Kimura), a position that some will dispute. A chapter on "Hormonal Influences on Extrapyramidal Sensorimotor Function and Hippocampal Plasticity" (J. B. Becker) documents hormonal effects on the structure and development of regions of the mammalian brain, but the specific mechanisms linking those changes to behaviour are not always clear. The text seems to be enviably free of errors, apart from a label inversion on a table of dissociation of gonadal activity from mating behaviour (p.157) that contradicts the text.

With such a broad mandate, it is more important for the book to be selective than to be comprehensive, but it is tempting to make suggestions for the next edition. New developments in biology at both ends of the reductionistic-holistic spectrum are ready to be incorporated. A new edition should cover more of the pharmacological and molecular approaches to behavioural endocrinology exemplified by the work of T. R. Insel, K. M. Kendrick, E. B. Keverne, B. S. McEwen and D. W. Pfaff, and the environmental and field endocrinology of physiologically minded behaviourists such as B. K. Follett and J. C. Wingfield, that begins to provide us with top-down explanations of how species differences in mating systems and breeding seasons evolve. More could be done to relate the material on aggression and circadian and circannual rhythms to species differences in natural history and life cycles. The truth is that, despite the exhortations of Beach, we have hardly begun to tap the rich source of new insights into the way in which hormones exert their pervasive effects on brain structure and behaviour that the comparative approach can provide. □

Peter Marler is in the Animal Communication Laboratory, University of California, Davis, California 95616-8761, USA.

Inheriting responsibility

Laurence D. Hurst

The Engineer in the Garden. By Colin Tudge. Jonathan Cape: 1993. Pp. 398. £17.99.

Gene Future. By Thomas F. Lee. Plenum Press: 1993. Pp. 339. \$24.95, £19.96.

PERHAPS more than any other discipline, genetic engineering figures in the public eye in an equivocal light. On the one hand, white-coated geneticists are caricatured as the devoted fighters of cancer and disease but, by equal measure, they could



Creators of monsters and marvels — in the eyes of the public, genetic engineers have the power to do enormous good but, by the same token, they also have powers of untold potential evil.

be the cruel experimenters on human embryos and the like. Jekyll and Hyde are alive, well and hard at work creating transgenic monsters and marvels. There is both excitement about what genetic engineering can do and concern about what it could do.

Some of this concern is perhaps more the result of confusion than consideration. Some is not. And hence any book explaining in a fair and accessible fashion just what genetic engineering can do and what the moral issues are is to be warmly received. Unfortunately, authors competent at explaining both the scientific facts and the ethical debate are rare, so two sorts of book tends to emerge: those concentrating on the ethics and those

concentrating on the genetics. *Gene Future* is in the latter class.

The structure of the book is very much of a form. The author starts by discussing DNA and inheritance, then moves on to explain what manipulations can be done, how they can be applied and what of good and bad may come out of it all. His treatment of the facts is always clear and rigorous. If anything, this rigour tends to degrade the text into a series of lists which seem out of place in a 'popular' book. These lists, however, serve Lee's overriding point, namely that the promised genetic revolution is upon us and we must now decide what we want to do with it.

As intimated, Lee does not go on to provide a guiding hand into the moral debate. Although he typically illuminates

the scientific facts around the domain of the ethical issues, the difference between germ line and somatic modification for instance, he does not lay out just what the alternative moral standpoints might be. Instead he describes the research bodies thinking about the issues and the policing authorities watching over the new genetics. Is this intended to reassure us?

One that it does not reassure is Colin Tudge. He argues that the policing agencies are devoid of moral authority. Lawyers are law enforcers, not moral advisers. As might be guessed from this strong attitude, the author is not satisfied simply to lay out the science. It is indeed the moral problems posed by the new genetics that inform much of Tudge's book. His is not, however, the book of an analytical philosopher dissecting the alternatives and deriving the consequences.

It is more an advocacy of a moral standpoint and is by turns invigorating and infuriating. Although most of the book is a very accessible introduction to evolutionary and molecular genetics, the author reveals his colours in the final few chapters. Here he puts genetic engineering in the context of the problem of technology as a whole. This in turn he merges with what he sees as the great problem facing us: ever-expanding populations moving us relentlessly towards the brink.

Whereas Lee realizes the same problem and then retreats, Tudge tackles the issue head on. His solution is twofold. First, he suggests a programme for public scientific literacy. Understanding must come before judgement. Citing Christ and John

The Ronald Grant Archive

Maynard Smith he also argues that "only the meek *can* inherit the earth": unless, in opposition to what game-theory tells us, we are all more selfless we shall go the way of all flesh. To impose the altruistic spirit, we must, he insists, reaffirm religious ideals and reinstate a priest class. Is the author arguing that we should all be religious or that a controlling class should manipulate the masses by encouraging religious ideals? Either way, from an understanding of the author's possible religious inclinations, many of his opinions start to fit into a mould. His appeal against mechanistic views of consciousness and of life, for instance, have a decidedly metaphysical ring to them.

Although I, and I suspect many others, can concur with many aspects of the author's ethical stance, for instance his concern for a moral standard in the treatment of animals, his affirmation of religion is to me unsatisfactory. Just as science got us into this mess, is not

science the best way out? But how then, one might argue, can we expect the necessary selfless acts of martyrs and saints? Do we not need a higher ideal to inform our morality? Perhaps so, but not one, I would argue, based on unquestioning obedience. During the siege of Lenin-grad in 1941–42, at least nine scientists, Lee reports, at the Vavilov Institute of Plant Diversity starved to death rather than eat the seeds that they had been entrusted to preserve. Although I find this story almost too incredible (and would like to know more), I would also like to believe that they died because of their understanding of the importance of biodiversity. If so, then it is perhaps to these few that we must look for inspiration, for, unlike any deism, their higher ideal was rooted in solid ground. □

Laurence D. Hurst is in the Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

these, I would have included the lurid account by Hector Berlioz of his first day as a medical student in one such charnel house. Eugene Sue was another who described these and other loathsome aspects of life in the Paris of the poor during the last century.

To list omissions is too easy of course, and probably otiose, for a good anthology has to reflect the tastes, literary and historical, of its editor and nobody else. All the same, while I found it hard to relate the extracts from Brillat-Savarin, Malthus, Jane Austen and one or two others to medicine (not of course that they are any the worse for that), I did rather miss A. J. Cronin, Francis Brett Young and Axel Munthe, all celebrated for their tales of the medical life in their time. No doubt Gordon felt the style to be too faded for the robust tastes of today. But then Gordon scores a wholly unexpected winner with a clutch of Victorian lady novelists, all of whom laid aside the stethoscope for the pen. The extracts, which could adorn a *Stuffed Owl* of prose writing, will leave Messrs Mills and Boon reeling.

Gordon's choice of verse includes broad swathes of W. E. Henley — surprisingly affecting on his long years as a patient — and of Robert Bridges, the doctor poet laureate (the subject incidentally of one of the great newspaper headlines, which appeared after he had failed on his arrival in New York to recite for the reporters who greeted him at the dockside: "King's Canary Won't Sing", it proclaimed). Gordon does not mention Bridges' account of life in the outpatient clinic at Bart's, where the average invalid received 1.28 minutes of attention, even if at death's door, and the casualty doctor 0.7 pence per examination. For my part, I think I should have included Thomas Hood ("I vowed that you should have my hand/But fate gives us denial/You'll find it there at Mr Bell's/In spirits in a phial" . . . and more in this vein), and perhaps a restorative dash of Belloc? Or among more modern and substantial figures, there are William Carlos Williams and Danny Abse, who both drew inspiration from their doctoring. But perhaps more typical was Sir John Hill, MD, playwright, who lives on in Garrick's celebrated epigram: "For physic and farces his equal there scarce is/His farces are physic; his physic a farce is." You will not find the hapless Hill in Gordon's pages, but there are riches in plenty and catharsis for all; and if you are of a nervous or hypochondriacal disposition, then take comfort from Napoleon's observation at a low point in his life: "Well, at least there is always death". □

Walter Gratzer is in the MRC Muscle and Cell Motility Unit, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK.

Strong medicine for weak stomachs

Walter Gratzer

The Literary Companion to Medicine. By Richard Gordon. *Sinclair-Stevenson*: 1993. Pp. 431. £18.

AN odd lot, doctors. To engage from choice in the exploration of other people's bodily orifices or shoving a fist into an abdomen full of quivering offal surely betokens an unusual cast of mind. Are they perhaps seeking to exorcise the daemons of death and disease, or is there a touch of necrophilia in some dark crevice of the psyche? Richard Gordon's excellent but unsettling anthology will do nothing to allay such suspicions, for it has much of the grisly fascination of the Hunterian Museum. It progresses from Sir Thomas Browne on the disposal of the dead — death here at its most deliquescent, all worms and snakes "out of the spinall marrow", luvacious liquors of the body and grave wax — through Pujol the Petomane, the musical virtuoso of the anal sphincter, to George Orwell, down and out among the paupers in a Paris ward of primaeval squalor, to conclude magnificently with Gordon's own finest creation, Sir Lancelot Spratt, the imperious surgeon in *Doctor in the House*.

Orwell's memoir is not easily eradicated from the memory. In Hôpital X (which Gordon identifies as the melancholy Hôpital Vaugirard) he looks on as death releases an old man from some hideous disease, only a foot or two away: a "natural" death, Orwell ruminates, such as you pray for in the Litany. "'Natural' death, almost by definition, means something slow, smelly and painful. Even at that it

makes a difference if you can achieve it in your own home and not in a public institution. This poor old wretch who had just flickered out like a candle-end was not even important enough to have anyone watching by his deathbed. He was merely a number, then a 'subject' for the students' scalpels . . ." Powerful stuff and not for weak stomachs. More Grand Guignol, this time unfamiliar, at least to me, is a description by the Hungarian humorist, Frigyes Karinthy, writing in 1935, of his operation for the removal of a brain tumour under local anaesthetic — the very stuff of nightmares. Gordon has a keen ear for commanding prose, and his chosen passages from Melville, Trollope, Flaubert (translated by himself) and Robert Louis Stevenson are guaranteed to bring up gooseflesh or make the gut heave.

There are gruesome narratives from fiction and fact of the deeds of the Resurrection Men, though Jeremiah Cruncher from *The Tale of Two Cities* does not appear, nor yet the reprobates who robbed the dead on the Napoleonic battlefields of their teeth to grace the dentures of well-heeled parties back home ("Waterloo teeth" they were called). He includes some fragments of verse by George Crabbe, but not the poet's story of how, when he was a medical student, his landlady found a dead child in a cupboard in his room, which she convinced herself was her own recently departed William. In France there were no Resurrectionists: the bodies of those who died in hospital and were not claimed within the day were rushed to the dissecting rooms. As to

A restricted view

Derek Fordham

Gender on Ice: American Ideologies of Polar Expeditions. By Lisa Bloom. University of Minnesota: 1993. Pp. 163. \$34.95 (hbk); \$14.94 (pbk).

Reading National Geographic. By Catherine Lutz and Jane Collins. University of Chicago Press: 1993. Pp. 309. \$59.95 (hbk); \$19.95 (pbk).

"COOK was a liar and a gentleman, Peary was neither", said Peter Freuchen, who knew them both in 1909 when they were squabbling over who, if either, had priority at the North Pole. He probably had a point because Peary was certainly reticent about the Eskimo mistress and several children he had left in the far north. His claim that "the pole is mine" showed similar amnesia about his black companion Henson who stood at his elbow and the Eskimos without whom he could not have reached wherever he was. Peary was indeed a uniquely awful man and it is such episodes that bring him into focus as the main character of Bloom's sincere but at times rather convoluted discourse. The author offers an analysis of the American ideology of polar exploration based on sex and race as the framework for a case study of Peary and the active role of the National Geographic Society (NGS) in promoting and defending him. However, the ease with which Peary found it possible to go his ungentelemanly way, his outrageous arrogance and high self-esteem made him a very difficult protégé.

These books are works of modern social anthropology, both centring on assessments of the NGS. Lutz and Collins go beyond Bloom's case study to examine the cultural issues behind the Society's glossy product — the *National Geographic*.

Both books deal with the rise of the NGS to power. It was power achieved in the main through its use of the latest advances in photographic techniques. It developed what came to be known as "the language of the photograph". This was a method of satisfying the requirements of the 50 per cent or so of readers who, we are told, explore the world of the *Geographic* no further than the sumptuous illustrations. By filtering fiction out of its publications it laid claim to the status of a scholarly national institution. It offered its readers only authentic material, claiming, in the words of its editor Wilbur Garrett, "to give the world what it doesn't know it wants yet". It began to vaunt itself as an instrument of government policy, not a mere publisher of a popular magazine. It was the ideal patron and image-maker for the likes of Peary. Peary became a national hero, not because of his personal qualities but as a result of the endorsement of his claims by the NGS.

As for race and sex, the America the NGS reflected in Peary's time was one in

which women scarcely featured and black women, with the luckless Henson, not at all, unless it was as a subject for one of the photographs of bare-breasted native maidens, the display of which was a practice the *Geographic* initiated well in advance of other publications. Such shots are seen by Bloom, Collins and Lutz as



On thin ice? — Peary claimed "the pole is mine".

obliquely emphasizing the essentially white male dominance of the magazine. Such dominance is reflected by Melvin Payne, one-time president of the NGS who, when faced with a photograph of a partially naked Polynesian girl of suspiciously fair complexion said: "We darkened her down to make her look more native, more valid you might say".

The absorbing and lengthy study of anthropologists Lutz and Collins examines the NGS largely through the processes by which the spectacular photographs for which the NGS remains world famous are acquired and used. They also show how "the window on the world and all that is in it" that the NGS claims to open to its readers affords only a restricted view, compatible with the mores of middle-class America and pro-government attitudes. "If the photo-

grapher didn't find any happy people I'd tell him to go back and find them", was an editorial directive that confirms the *Geographic* as an instrument of manipulative mass culture. Objectivity comes a poor second to editor Gilbert Grosvenor's 'seven principles' for an acceptable *Geographic* piece. Principle six specifies that, "only what is of a kindly nature is printed". Set out in 1915, they are still reflected in the sanitized reporting of any potentially uncomfortable issue — and this is the third most popular magazine in the United States.

The complex images of the NGS, set out and criticized by Bloom, are enlarged on by Lutz and Collins who provide a fascinating insight into the intricacies of the photographic processes of the *Geographic*. They see the NGS's use of photography as symptomatic of an institution purporting to promote humanity and understanding while holding back truly accurate appreciation of non-Western peoples and controversial issues. All this in the interests of a balanced or 'kindly' image. Grosvenor's rules of 1915 still apply — nowhere more so than on the occasion on which a loincloth did not provide quite enough coverage for the *Geographic's* sensibility and "shorts to be added by engraver" was pencilled on the proofs.

The *Geographic*, as these books clearly reveal, often leaves out more than it contains. It is as uniquely American a phenomenon as Peary and apple-pie. It seems likely that few of its many European readers will have perceived the conflicts simmering between the familiar yellow covers that Bloom, Collins and Lutz expose to view. Many will probably find these days that their thoughts on its arrival turn to reassessing the structural

implications of the ever-growing stack of those not-to-be-discarded copies in the attic.

For *Geographic* readers with space in the attic, these books offer stimulating, sometimes over-long insights into the cultural background and decision-making of what, as they reveal, has become a national institution. "Geographical societies would turn into ineffectual adventure clubs without substantial links to science", P. Pauly recently reflected on the position in 1903. For the NGS today, poised on the divide between science and entertainment, things have not changed much. □

Derek Fordham is secretary of the Arctic Club, and is at Arctic Camera, 66 Ashburnham Grove, Greenwich, London SE10 8UJ, UK.

Extrasensory deceptions

Steve Blinkhorn

Parapsychology: A Concise History. By John Beloff. *Athlone*: 1993. Pp. 330. £32 (hbk); £12.95 (pbk).

SOME things never change. Sixty years ago, Harry Price published *Leaves from a Psychist's Case Book*, a compendium of some of his more interesting forays into purported paranormal phenomena. It is characterized by the excessive use of italics for emphasis and exclamation marks for assertion when the evidence is thin or the author feels his argument is less than convincing. It also contains some extraordinary photographs, which nowadays have the kind of period feel one sometimes associates with antique surgical apparatus. I am rather attached to my copy, which I inherited from my grandmother-in-law, partly because she squandered her modest fortune on supporting Harry Price's activities.

Sixty years on, Beloff's history of parapsychology has its full quota of italics and exclamation marks, lacking only the photographs. This is a shame, because its author seems to believe that thoughtography — the creation of photographic images on film by thought alone — is a genuine phenomenon. You can put photographs in a book, whereas ectoplasm eludes the publishing process. Photographs of ectoplasm might be a good substitute, for as Beloff puts it, with regard to ectoplasmic rods supposedly responsible for the levitation of a table: "Like all photographs of ectoplasm, these curious structures may look spurious . . ." But we have to take his word for it, as all he gives us is text.

Ectoplasm, which when photographed tends to bear a striking resemblance to cheesecloth or butter muslin, generally has a marked preference for issuing from the upper rather than the lower end of the gastrointestinal tract (although in the case of the medium concerned in this particular case it came "mainly from her vagina"). Ectoplasmic manifestations have declined markedly in recent decades. I attribute this to the decline in the availability of

cheesecloth and butter muslin as a consequence of modern methods for the bulk processing of dairy products.

In fact, strong paranormal phenomena of all kinds have declined almost to vanishing point, just as good scientific methods for studying them have reached perfection — infrared cameras for the materialization of spirits in seances, for instance. Why should this be? I have three solutions to offer. First, a lot more people have died since the 1930s, and the consequent congestion in communication channels between this life and the 'other side' prevents effective and reliable contact. This we might call the cellular telephone



Two heads are better than one — "Like all photographs of ectoplasm, these curious structures may look spurious . . ." The medium Marthe Berauld tested by Schrenke-Nötzing, from a photograph in the Harry Price Collection.

hypothesis. Second, we have the politically correct hypothesis: Red Indian spirit guides have withdrawn their labour until the history of their participation in psychical research is rewritten, and we start calling them Native Americans as is their due. You can probably guess the third.

There has even been a shortage in recent years of spoon-bending phenomena, which attracted so much enthusiasm in the 1970s. Could it be that the widespread availability of video recorders destroys the conditions of calm and concentration upon which they depend?

"The community of sceptics is always ready to scoff", would sum up Beloff's response to such suggestions. But Sir

Oliver Lodge was quite keen on the idea that thought transference could be seen as an electromagnetic phenomenon transmitted through his beloved ether, and I fail to see why paranormal phenomena should not suffer the same restrictions as my mobile telephone. Anyway, as a hypothesis it is no more daft than his own suggestion, that nature acts as an extraordinary sort of immune system, fighting off paranormal manifestations, presumably with antispirits rather than antibodies.

Beloff repeatedly endorses the argument that where no conventional explanation of supposed paranormal phenomena has been forthcoming this is because no conventional explanation is possible. This is the common fallacy of paranormal claims. Similarly, in quoting the odds against particular results arising by chance, he takes the improbability of chance as suggestive of the probability that something paranormal has occurred. This style of argument will be depressingly familiar to many journal editors, not to mention those charged with instilling the rudiments of hypothesis testing in unwilling undergraduates.

Now there is one curious feature of nearly all of the cases he discusses: sooner or later the medium is discovered cheating, or poor experimental control is demonstrated. Psychic phenomena decline when investigators start to work out how they might be achieved by normal means. If we are to take Beloff seriously then providence prefers potential cheats when determining how to distribute psychic powers, as he is prepared to believe in the reality of the phenomena produced on those occasions when the actual method of cheating was not exposed.

This is a concise history, made all the more concise by the omission or minimization of cases that once made headlines, but subsequently proved insubstantial. For instance, who can forget Carl Sargent's experimental work at Cambridge in the 1970s? Here he merits merely a mention in the notes and bibliography. We are left with the flotsam and jetsam of not quite totally discredited phenomena to interest true believers, a lot of frank description of fraud, some interesting pottered early history, and a distillation of post-Renaissance philosophy of mind and miracles, which suffers greatly from a failure to consider mediaeval scholastic views on the nature of the soul and the ordering of nature.

Perhaps in sixty years' time this book will seem as quaintly outdated as Price's *Leaves*. Or perhaps it won't take nearly as long as that. □

Steve Blinkhorn is at Psychometric Research and Development Ltd, Brewmaster House, The Maltings, St Albans, Herts AL1 3HT, UK.

Thinker's progress

Richard Davenport-Hines

Paradise Dreamed: How Utopian Thinkers Have Changed the Modern World.

By Pamela Neville-Sington and David Sington. *Bloomsbury*: 1993. Pp. 322. £18.99.

"PROGRESS", said Oscar Wilde, "is the realization of utopias": an opinion with which the authors of this cultural history of utopian thinking generally agree. They cover their subject with the soothing syrup of twentieth-century liberalism, and conclude that "many of the most positive political developments of the past half-millennium have also drawn nourishment from the utopian tradition: the rise of science, of female emancipation, of democracy itself". This conclusion is made possible by their own optimistic spirit and by a certain selectiveness in their material; it is not the result of ideological dishonesty.

Paradise Dreamed is most satisfying when dealing with quasi-scientific topics. The book begins with a sketchy introduction to Plato and the Golden Age of ancient Greece and a rather unexciting summary of Thomas More's *Utopia* (1516); but it becomes more interesting when discussing the utopian speculations of Francis Bacon. This great English jurist, who regarded philosophy as having no practical use, propounded a new role for science in society in his *New Atlantis* (1627). Systematic experiment and scientific innovation would achieve "the enlarging of the bounds of Human Empire, to the effecting of all things possible", Bacon hoped. To this end he proposed the foundation of a research laboratory to be named the College of Six Days' Works which was to include cave-houses for experiments in refrigeration, astronomical towers, desalination plants, ecological mini-environments, furnaces, acoustic chambers, engine-houses of all sorts and much else including "houses of deceits of the senses". Readers who are unfamiliar with Bacon's vision of a science-based utopia will welcome the Singtons' introduction to his ideas.

Lewis Mumford's view, that cities were the earliest utopias, is taken up by the Singtons, who give a plain and unpretentious presentation of European architectural ideas since the renaissance as expressions of utopian thought. Their section on urban engineering concludes with an account of the career of Le Corbusier which is so muted in its judgments that it will satisfy neither his opponents, who complain that he exemplified the most inhumane and totalitarian elements of twentieth-century architecture, nor his surviving protagonists. *Paradise*

Dreamed had its origins in a BBC radio series: unfortunately the need for 'broadcasting balance' has left a somewhat lifeless version of the authors' opinions.

Another section examines the impact of Edward Bellamy's *Looking Backwards* (1888), described as the best-selling novel in the nineteenth-century United States after Harriet Beecher Stowe's *Uncle Tom's Cabin*. Bellamy's fantasy about the benefits of mechanization prefigured the glories of a consumer's paradise, launching a literary tradition in support of technological progress that was halted only by Aldous Huxley's *Brave New World* (1932). Both books, as analysed by the Singtons, contain ideas about applied science that are thought-provoking if anecdotal. The difficulty of sustaining a long analysis of serious ideas in a radio programme is discernible even in this discussion of Bellamy and Huxley.

There are serious issues underlying the chatty approach. "I deny that freedom exists at all", declared the behavioural psychologist B. F. Skinner, whose novel *Walden Two* (1948) is analysed by the Singtons. "I must deny it — or my program would be absurd." Skinner was one of the few modern utopians to recognize explicitly that freedom and organization are incompatible, and that effective planning has to be oppressive. Other malign features of utopian thought are identified by the Singtons. "We must build on a clear site", said Le Corbusier, expressing the utopians' instinct to reject history and erase the past. Each utopian scheme is the projection of a long imagination, and seeks to impose the values of one individual on a community. Non-conformity can become a criminal transgression, and there is a tendency to seek equality by enforcing uniformity. Although Marx abhorred utopianism as reactionary, communism is rightly treated by the Singtons as a potential paradise perverted by evil.

"What has always made the state a hell on earth has been precisely that man tried to make it his heaven", wrote F. Hoelderlin. Schemes of human perfection prove in the end only human imperfectibility. One does not have to be a Christian believer to regret the Singtons' concentration on Western secular thinkers; it means that although their book is essentially a reading of Western literary texts on ethical themes, they do not mention several great ethical writers who put their ideas into a Christian framework without evangelical attitudinizing. Foremost among these in the twentieth century is W. H. Auden. Unlike the melioristic, materialistic Singtons, Auden, after long thought about both arcadian and utopian visions, concluded that "our world rapidly worsens". His great poem *Vespers* (1954) is one of the sharpest ever critiques of the utopian mentality. At first it seems frivolous in its teasing of the *dirigistes*, technocrats and

zealots of utopian reform; but he offers an alternative vision, and ends by adjuring that "without a cement of blood (it must be human, it must be innocent) no secular wall will safely stand".

Although the Singtons are sympathetic to communitarian experiments such as the Shakers in the United States, and see attractions even in Californian hippydom of the 1960s, their optimism is not convincing. "The attempt to externalize the kingdom of heaven in a temporal shape must end in disaster", wrote Auden's contemporary Hugh Kingsmill. "It cannot be created by charters or constitutions not established by arms. Those who seek for it alone will reach it together, and those who seek it in company will perish by themselves." □

Richard Davenport-Hines is at 51 Elsham Road, London W14 8HD, UK.

Bugbear

Arno A. Penzias

Digital Woes: Why We Should Not Depend on Software. By Lauren Ruth Wiener. *Addison-Wesley*: 1993. Pp. 245. \$22.95, £19.95.

ANYONE who still regards computers and the software that controls them as beyond blemish will find readable antidotes in this volume's case histories. A veritable 'Murphy's Law' (if anything can go wrong, it will) for software, the presentation will reward readers who seek cause to worry about the programs that 'chug' away on their behalf. Although computer scientists will find themselves hard put to accept many of the sweeping statements that pepper the text (such as "the software just can't keep up [with hardware advances]" and "testing software thoroughly is simple — and impossible"), the worst one can say about the author is that she is unnecessarily pessimistic about prospects for quality improvements. So fine, let's try and prove her wrong.

The book begins with 13 accounts of technology-related mishaps, and ends with 7 criteria for selecting a software system. In between, readers receive a lively — if rather downbeat — view of the software industry as might emerge from an after-work gripe session. Except for Bill Gates ("a CEO [chief executive officer] who can program"), industry executives rarely do the right thing, unless tricked by resourceful subordinates. True to the subtitle, the presentation attempts to persuade readers not to rely on software. But quantitative data that might test this thesis are given no place in this anecdotal format.

How has the growing use of software

affected the quality of the systems we moderns depend on? Have aeroplanes, telephone networks, power plants, department-store bills and the like become more reliable or less with the spread of microchips? Answers to such questions lie beyond the scope of this book.

Instead, Wiener examines each advance for potential threats. With repetition, this litany wears thinner as it proceeds. Publishing books electronically will lead to multiple rewrites of the classics. "How are we going to tell which is the original?" the author asks, thereby rekindling a debate that must have accompanied the advent of movable type.

A defender of science

A. Rupert Hall

Defining Science: William Whewell, Natural Knowledge and Public Debate in Early Victorian Britain. By Richard R. Yeo. Cambridge University Press: 1993. Pp. 272. £35, \$54.95.

WILLIAM Whewell (1794–1866), the pivot of this study, a close contemporary of Babbage, Herschel and Peacock, was a greater pundit than any of these. Ascending the successive steps of classicist, mathematician, priest and professor of moral sciences, he at last attained one of the finest offices in the British prime minister's gift, the mastership of Trinity College, Cambridge: a reward not often given to a Fellow who has made his way up the college ladder. Yet, as Whewell recognized regretfully, no theorem or discovery bears his name. His early textbooks of mechanics were good, he contributed to the modernization of Cambridge mathematics and mineralogy, he stimulated and advanced global knowledge of the tides; but even these relatively modest accomplishments were completed before he was much over 40. The rest of his work and the bulk of his fame lay outside scientific teaching and innovation.

If this was Whewell's science — which is not of prime concern to Yeo — what was his "omniscience", his "foible" according to Sydney Smith? As for so many Victorians, Germany was an intellectual inspiration: Whewell published a book on Gothic German architecture, trans-

lated German poetry, and himself wrote English hexameters; he read the German moralists — not, perhaps, the German theologians who so impressed the young George Eliot. He preached influential sermons, he wrote books on morality, he contributed to the mid-Victorian educational debate (ever advising that the Cambridge disciplines of Greek and geometry were the essential foundations of learning), he deplored the Oxford Tractarians and defended the divine creation. Above all, Whewell wrote universally about science — Yeo would amend Sydney Smith's quip to make "metascience" Whewell's foible. Some of this bulky writing was addressed to readers of the reviews — perhaps one would today hardly imagine Sir Michael Atiyah composing an essay equivalent to *The General Bearing of the Great Exhibition on the Progress of Art and Science* (1851) — much was at the highest level of philosophical debate, notably in his *Philosophy of the Inductive Sciences* (1840). His dissensions from Locke, Comte, J. S. Mill and other giants remain to this day of lively technical interest. Nevertheless, one sees what a vast intellectual gulf divides Whewell from such a man of the next generation as James Clerk Maxwell, not to mention Charles Darwin, 15 years his junior and a practitioner of sciences still regarded by Whewell as speculative.

Arno A. Penzias is vice president of research at AT&T Bell Laboratory, PO Box 636, 600 Mountain Avenue, Murray Hill, New Jersey 07974-0636, USA.

What did Whewell hope for from his "metascientific" studies, publications of deep thought and intense effort? He was far from alone in embarking on a 'public debate' about 'natural knowledge in early Victorian Britain' (to rearrange Yeo's subtitle): Babbage, Herschel, Brewster, Sedgwick, Murchison and many others aired their views on what science is (or is not), what its boundaries and prospects are, why it is necessary to mankind and why it should be promoted (or discouraged). Like many contemporaries,

Whewell held the most advanced and mathematical departments of science, such as astronomy and mechanics, to be essentially perfect and infallible; one problem therefore was the relationship of contemporary non-mathematical sciences, and mathematical science before Newton, to these perfected bodies of knowledge. Whewell discussed the first of these aspects of the problem in his *Philosophy of the Inductive Sciences*, the second aspect in his earlier *History of the Inductive Sciences* (1837). These two closely connected books are the twin props of his reputation today.

Whewell was a massive and authoritative figure. Perhaps his philosophical weight was most effectively cast against a facile Baconianism or positivism, and he may have agreed with a friend that 'Plato is worth 10,000 Aristotles and 100,000 Lockes', for Locke's philosophy treated the human mind as if it were "the same sort of thing as Babbage's Calculating Machine" — a surprising comparison for 1822. It was a philosophy (Whewell thought) hostile to both science and religion.

Yeo writes of Whewell as a defender of science, though he never clearly defines the opposition. Whewell was never so hot a critic of governmental apathy and the Royal Society as was Babbage, nor did he believe that universities should be seats of research. He argued that utilitarianism misrepresented science, without alleging that this philosophy obstructed it. Classicists might maintain the superiority of their studies for human life, but with them Whewell had no dispute. High churchmen (like popes) distrusted science; Whewell (like Newton) insisted on the reconciliation of religion and science. Essentially his ideas were conservative (arousing Huxley's criticism) as may be seen in his rejection of Lyell's geology (in the concluding pages of his *History*) and the whole final chapter of *On the Philosophy of Discovery* (1856).

A more consistent concentration on Whewell's (qualified) role as an advocate of scientific knowledge, and of the importance of the endeavour to understand nature within European thought, might have brought a sharper focus to this rather diffuse study, which still leaves Whewell's stature enigmatic. The importance of the great reviews in British intellectual life in his time is well brought out; the *History* is well assessed; the treatment of Whewell's philosophy is heavy going. Fashionable use of such words as 'space', 'visibility', 'agenda', 'rhetoric' and 'discourse' in a figurative sense does not make for clarity. A book entitled *Defining Science* might reasonably have ended with a statement of Whewell's definition. □

A. Rupert Hall is at 14 Ball Lane, Tackley, Oxfordshire OX5 3AG, UK.

1994 review supplements

Nature's review supplements next year are Spring Books (24 March), Autumn Books (17 November) and New Journals (29 September). The latter will cover journals launched during or after June 1992 with four separate numbers issued by the end of April 1994.

The enemy is us

Philip Shabecoff

Global Accord: Environmental Challenges and International Responses. Edited by Nazli Choucri. MIT Press: 1993. Pp. 562. \$50, £44.95.

Ultimate Security: The Environmental Basis of Political Stability. By Norman Myers. W. W. Norton: 1993. Pp. 308. \$25.

Threats Without Enemies: Facing Environmental Insecurity. Edited by Gwyn Prins. Earthscan: 1993. Pp. 197. £12.95.

Global Political Ecology: The Crisis in Economy and Government. By Adam Swift. Pluto: 1993. Pp. 345. £35 (hbk); £14.95 (pbk).

THE interdependence of living organisms and their habitat is the first principle of the science of ecology. It is also a sacred article of faith of modern environmentalism. One of the most frequently quoted pronouncements of John Muir, the gaunt naturalist and mystic who is a patron saint of the American preservationist movement, is that "When we try to pick out anything by itself, we find it hitched to everything else in the universe."

The 'universe' Muir was talking about is the biological universe of living plants and animals and the resources that sustain them. Over the course of this century, the environmental movement has gradually emerged around the world as a response to the way in which human activity has degraded, and in some cases begun to destroy, the natural world that sustains life, including human life. That response has intensified with the realization that the degradation and destruction are starting to reach planetary proportions.

For decades, the environmental response focused on the linear effects of human activity on ecosystems, including such problems as population pressures, industrial pollution, the eradication of wildlife and its habitat, the erosion of land, the flood of hazardous substances into the environment, excessive use of fossil fuel energy, nuclear hazards and the general impact of applied science and technology on human health and the natural world.

In recent years, however, it has slowly dawned on environmentalists and policy-makers alike that the fate of

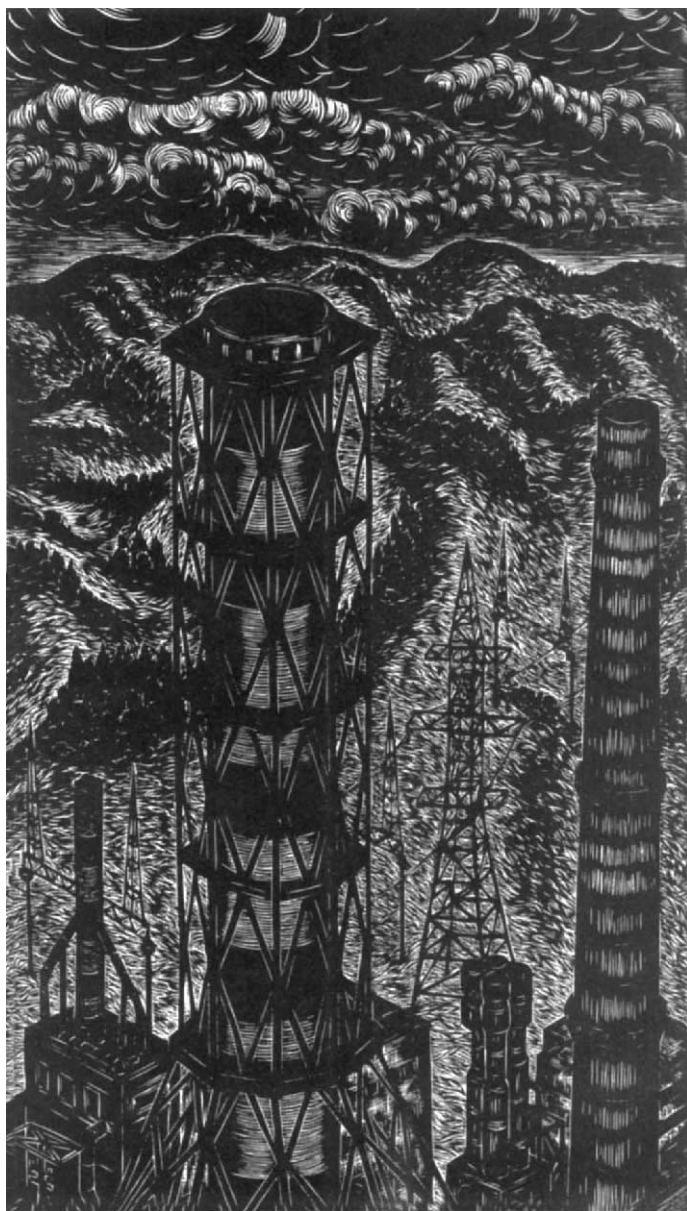
the community of life is hitched not only to overt human assaults on the environment but also to human systems and institutions — economic arrangements, political gov-

ernance, social and legal structures and, increasingly, relationships between and among nations. There has also been a growing awareness that the way in which these relationships are addressed by the international community will do much to determine whether there will be a workable system of collective security in the post-Cold War, post-industrial era. That realization was the underlying agenda of the 1992 Earth Summit in Rio de Janeiro. And it is the underlying theme of four new, very different books published in the United States and the United Kingdom.

The linkages between the created world of human institutions and systems and the fate of the global environment are addressed exhaustively in the most formidable and potentially the richest of the four

books, *Global Accord*, edited by Nazli Choucri, a political scientist at MIT. A chief goal of the book, Choucri says in an introductory chapter, is "relating environmental variables and process to social activities, national characteristics and international relations". This goal is pursued in 15 chapters by academics from a variety of disciplines, including political science, law, economics and ecology. The subjects include a review of the physical threats to the planet, the flaws in our technological, economic and political systems that are responsible for those threats, the intellectual transformation in thinking about the environment, and detailed examinations of how institutions and legal structures might be altered to respond to planet-wide crises.

All the essays are dense with information and several contain provocative arguments. One particularly compelling chapter on inter-generational equity, by Edith Brown Weiss, a law professor at Georgetown University, finds precedent in African, Asian and Islamic tradition as well as in Western law for requiring a commitment to preserving the Earth's health and resources for the benefit of future generations. A chapter by David G. Victor, Abram Chayes and Eugene B. Skolnikoff suggests parameters for building regimes to carry out international accords such as the Climate Change Treaty signed in Rio. These would include flexibility, transparency,



Michael Dalcero

War for the world — response is now focused less on the linear effects of human activity and overt assaults on the environment, such as nuclear hazards (including Chernobyl illustrated above), and more on the effects of human systems and institutions.

decentralization of monitoring and enforcement, fairness in sharing the burden, cooperation rather than intrusive implementation and a "significant role" for non-governmental organizations.

Global Accord is a scholarly book in the best and worst senses. At its best it is a comprehensive look at the incredibly complex set of issues that are encompassed by the concepts of global environment and sustainable development. These issues are examined in systematic, thoroughly documented fashion with occasional flashes of original insight. At their worst, some of the essays are virtually unreadable exercises in academic babble that bury the obvious under a stultifying layer of jargon. But for those willing to take the effort to dig hard, there are throughout this volume rich veins of data and bases for addressing what are arguably the most compelling problems facing the community of nations at the end of the twentieth century.

A very different sort of book is *Ultimate Security* by Norman Myers, an environmental scientist and combat veteran of the ecological wars. Myers, one of the early voices raised to warn, in books such as *The Primary Source* (Norton: 1992, 2nd edn), of the destruction of tropical forests and loss of biological diversity takes a personal, engaged view of the threats to global environment, their causes, their significance and their solutions.

The 'ultimate security' of the book's title is, of course, environmental security. "Security concerns", he writes, "can no longer be confined to traditional ideas of soldiers and tanks, bombs and missiles. Increasingly they include the environmental resources that underpin our material welfare. These resources include soil, water, forests, and climates, all prime components of a nation's environmental foundations." In fact, Myers finds that spending on armaments is actually undermining national and collective security by draining resources from programmes needed to preserve environmental stability. For example, he calculates that to provide safe drinking water to the one-third of the world population who do not have it now would take the equivalent of 12 days of current global spending on the military.

Ultimate Security, and indeed all four of these books, finds that demographic pressures as well as poverty within nations and economic inequity among them are major threats to global peace and stability. Myers devotes several chapters to describing how specific environmental problems, including disputes over scarce water in the Middle East, deforestation in the Philippines and land degradation in Mexico, are disturbing international and domestic tranquility.

Although he describes the quest for security as a task for all nations and all

individuals, Myers assigns special responsibility to the United States to show the way because, he says, it is now the only country with the technical and military power to do so and because it is also the world's biggest polluter. Although that is true, he is letting Western Europe, Japan and corrupt regimes in the developing countries off the hook.

But how can one not like a scientist who writes like this: "And the winds carry no passports. No more than rivers or birds, which likewise tell us that human-made divisions of our world are too artificial for the Earth under intolerable strain?"

The very notion of ultimate security, however, is dismissed by Sir Crispin Tickell, the veteran British diplomat who is now warden of Green College, Oxford. Writing in *Threats Without Enemies*, Tickell asserts that "There is, of course, no such thing as security, global or otherwise." Environmental insecurity is and will remain inevitable, he argues nevertheless that security is a goal that must be pursued. Just as peace-keeping is rarely undertaken unless there is war, there is not likely to be meaningful progress toward environmental security in the absence of environmental catastrophe, Tickell believes.

This kind of cold-eyed *realpolitik* can be found in most of the analysis in *Threats Without Enemies*, yet another set of essays, this one edited and coauthored by Gwyn Prins, director of the Global Security Programme at the University of Cambridge. In fact, one of the contributing authors, Hans Peter Durr, director of the Max Planck Institute in Munich, praises the "virtues of pessimism" in dealing with the complex issues of environmental security, a notion he finds "utopian". But dealing with the problems, he argues, will draw on utopian visions such as replacing our current "arrogant economy" with an economy founded on the principle of sustainable development.

But if *realpolitik*, defined as 'cynical disregard of moral considerations in political behaviour', is useful in analysis, it must be replaced by 'realism' in the political response to environmental insecurity, Prins contends.

One of the more interesting essays in this volume is by Admiral Sir Julian Oswald, First Sea Lord of the Royal Navy, who believes there should and will be a growing role for the military in protecting the global environment. The armed forces, he says, are well equipped to do much of the research into environmental threats and solutions and to monitor environmental trends. More surprisingly, he believes that the military, especially naval forces, will have a coercive role to play in enforcing international environmental agreements such as fisheries protection treaties.

It must be said, however, that *Threats Without Enemies* is not an appropriate title for what this book actually demonstrates. In fact, it points to a very real and relentless foe best described by the immortal words of Walt Kelly's Pogo: "We have met the enemy and he is us."

Finally, what to make of *Global Political Ecology*? Its author, described as a former civil servant and senior official, has written under the pseudonym Adam Swift. Because the book has much to say about free market economics, the 'Adam' is obviously borrowed from Adam Smith. And because he makes the modest proposal of scrapping our current economic and political systems, the 'Swift' is quite clearly Jonathan.

The author has constructed a kind of unified field theory to explain most of the world's woes, including environmental degradation, waste of resources, poverty in the developing world, recession, crime, drugs and, in the rich countries, terrorism, the abuse of human rights, the arms race and the threat of nuclear annihilation. All these problems and more, he argues with great intensity, spring from an economic system — and a political structure that supports it — designed more than 200 years ago for a world that has since changed beyond recognition. The "problem of problems", he contends, is that the international economic framework, based on greed and competition, "has reached what seems a definitive state of unworkability".

Swift marshals a staggering array of statistical data and other evidence to support his position. His text is accompanied by graphs, tables and flow charts so complex and bewildering that they seem to have been devised by systems engineers under the influence of a powerful hallucinogen. His proposals range from the utopian (a "broad balance of income security and distribution") to the orwellian ("citizens should be conditioned to realise the duties as well as the rights involved in social existence"). His goal, which is nothing less than to "civilise market practice" and approach "true economic democracy", is wildly beyond the reach of our late twentieth century civilization.

And yet, there is no denying the grim logic of Swift's analysis of the bleak landscape to which our current road is leading us. At the least, *Global Political Ecology* should provide a needed radical leavening to the active but somewhat bland current geopolitical debate over sustainable development.

All the books discussed here, in fact, are welcome additions to what is becoming a kind of global town meeting over the future of the planet. □

Philip Shabecoff is publisher of *Greenwire*, a daily environmental news service, Falls Church, Virginia 22046, USA.

The fairer science

Roy Porter

Nature's Body: Gender in the Making of Modern Science. By Londa Schiebinger. Beacon: 1993. Pp. 289. \$25.

THE rise of modern science meant the discrediting of the old human-centred subjective cosmos beloved of mediaeval people, which attributed soul and feelings to everything. Its gushing, poetic anthropomorphism was replaced by a worldview that was severely objective: nature was just matter in motion, governed by mathematical laws, studied by neutral scientists. Or so, at least, runs the conventional textbook story of the scientific revolution. But such readings are far too simple, argues Londa Schiebinger in her crisp and well documented account of Linnaean natural history and early anthropology. Science did not suddenly become value-neutral: it long continued to incorporate ideas that saw the kingdoms of nature through human categories (as the metaphor of 'kingdom' itself reveals).

"Leaves . . . serve as bridal beds . . . perfumed with so many soft scents that the bridegroom with his bride might there celebrate their nuptials with so much the greater solemnity." Who is writing here? Some bard waxing lyrical about a tryst in an arbour? No, it's none other than Linnaeus, the eighteenth-century founder of modern taxonomy, proudly displaying his understanding of recently discovered plant sexuality and describing the fertilization of flowers. Ah, but that's just a rhetorical flourish, it might be objected. Not so, however. For the truth is that Linnaean botany is saturated with sexual allusions and judgements more appropriate to human ethics than to nature. Linnaeus made the key to his sexual system what he termed the *nuptiae plantarum* — the marriages of plants. One class of plants (Linnaeus's *monandria*) practised monogamy. Other plants had twenty or more 'husbands' (Linnaeus's term) sharing the marriage bed (petals of the same flower). In diverse arrangements of stamens and pistils, plant husbands of the class significantly labelled *polygamia* lived with 'wives' and 'harlots' or 'concubines'. The writings of Linnaeus, and his British popularizers such as Erasmus Darwin, systematically project onto botany the idioms of human eroticism and sexual mores — as outraged moralists protested.

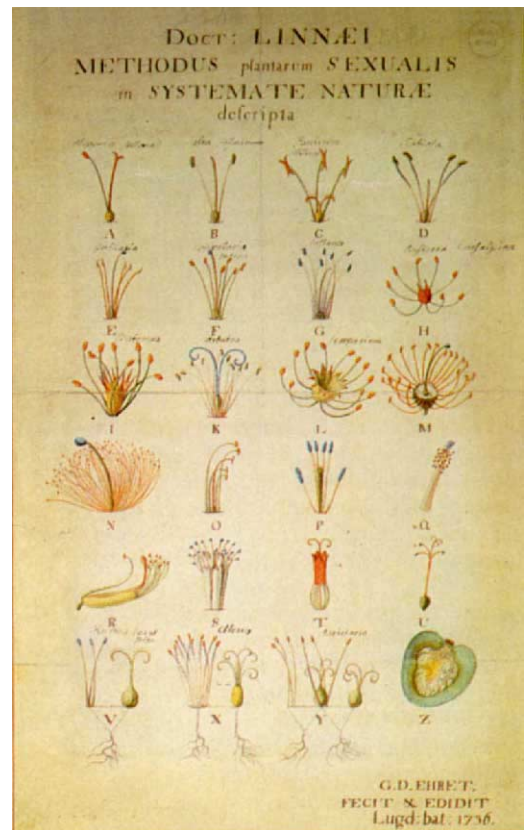
And nowhere is this so conspicuous as in another Linnaean coinage: mammals. Classification after Aristotle had distinguished 'higher creatures' from fish, amphibians and birds by calling them 'quadrupeds'. But naturalists grew aware that four-footedness was an unsatisfactory descriptor, for whales — and humans —

obviously shared such 'higher' features as warm-bloodedness, the four-chambered heart and hairiness. But in the end the distinguishing feature that clinched the classification Linnaeus devised was the possession of breasts (*mammæ*). Why? Not because there was no other possibility, Schiebinger reasonably argues — we could all be called *Pilosa* (the hairy ones); but in part because the idea of mothers suckling their young achieved tremendous kudos within human culture in the enlightenment era. Rejecting the prevalent fashion of wet-nursing, Linnaeus made his wife put all their seven children to the breast.

In discussing binomial botany and the naming of *Mammalia*, Schiebinger has no quarrel with the science produced. Her aim is merely to insist that to grasp why certain scientific ideals jell at certain times, it's not enough to think in terms of 'discoveries': we must have a view to the outlooks, interests — and sex — of the naturalists involved. Several of the chapters of her wide-ranging book, however, focus on areas of inquiry where science lent its authority to gross bias or bigotry. One was the creation of racial stereotypes.

An age of exploration produced greater awareness of 'primates' (another Linnaean coinage) and of the diversities of mankind. Not surprisingly, strenuous efforts were made to construct typologies of difference. Inevitably this entailed uncertainties about the relationships of "people of color" (Schiebinger's phrase) to gorillas, chimpanzees and the various 'pongos' and 'jockos' that figured in the travel literature. The superiority of 'Caucasians' (another new coinage) might be reinforced by the downgrading of blacks towards the apes — a process aided through the rather whimsical tactic of attributing a surprisingly high degree of 'savvy' and civilization to orangutans. In Thomas Peacock's satirical novel, *Melincourt*, Sir Oran Haut-ton is an orangutan who rises in society and (thanks to his silence) becomes a highly esteemed member of parliament.

Alongside 'people of color', early anthropologists stigmatized women, often grotesquely. Black women were often represented, in travellers' tales, as copulating with gorillas (hence miscegenation). Their supposed sagging pendulous breasts became objects of disdain, derision and prurience for male naturalists, as did the notorious elongated labia minora of Hottentot women. Once the so-called 'Hottentot Venus' was brought to Europe early in the nineteenth century, it seemed as if the whole community of naturalists



Plant sexual systems drawn by Linnaeus, the eighteenth-century founder of modern taxonomy.

was gawping, purely out of love of science, at the genitalia of black women.

A certain (perfectly justifiable) indignation informs Schiebinger's book. Science has much to answer for in its creation of successions of stigmatizing stereotypes, supposedly 'natural' but in reality the projection of prejudice. But she is no less sensitive to the complexities and the genuine dilemmas of scientific classification. Science (as the case of Sir Oran shows) created positive as well as negative images; blacks and women also had their defenders, and a few naturalists, such as Blumenbach, were deeply circumspect about drawing premature conclusions about the map of mankind. Moreover, science also punctured certain traditional myths, not least when Linnaeus actually set *Homo sapiens* among the primates.

Seeing nature through social spectacles remains a problem today, as ethnic and sex differences continue to be politically explosive and sociobiology creates new anthropomorphic languages — with talk, for instance, of genetic investment strategies. Schiebinger's incisive book draws timely attention to the critical roles of language, metaphor and ideology in the development of science. Anthropomorphization was never eradicated: it is our job to be alert to it. □

Roy Porter is in the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

Prehistoric archaeology

Martin Rudwick

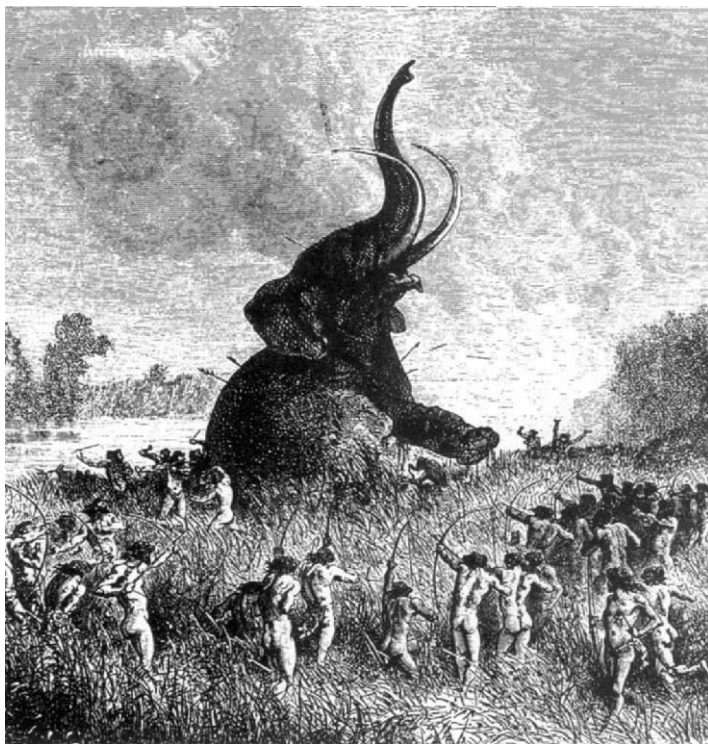
Men among the Mammoths: Victorian Science and the Discovery of Human Prehistory. By A. Bowdoin Van Riper. University of Chicago Press: 1993. Pp. 267. \$51.75, £35.95 (hbk); \$19.50, £13.50 (pbk).

THE year 1859 was an *annus mirabilis* in two distinct areas of science, a date almost worthy of 1066 and *All That*. The appearance of Darwin's *Origin of Species* showed publicly that a reputable naturalist had joined what had previously been the scientifically suspect side of the long-running debate about organic evolution. But the heightening of that debate, with its implications for human self-understanding, tends to overshadow events of equal significance a few months earlier, when a series of addresses given by certain leading geologists to some of the most respected scientific bodies in Britain presented a tightly argued case for concluding that early human beings had coexisted with the extinct mammals of the Pleistocene period. *Men among the Mammoths* is a highly readable account of the research that expanded the scale of antiquity for the human species, and integrated human history into the far longer narrative of Earth history.

This was indeed, a "great and sudden revolution", as one contemporary called it. In little more than 18 months a long-standing but dubious speculation was transformed into a solid consensus, at least among the scientists most competent to judge the evidence. Van Riper, however, rightly sets that exciting period into a longer and richer historical and scientific context. He shows, first, why the context of the new discoveries was geological rather than archaeological. In the early nineteenth century, archaeologists were mainly concerned with supplementing the documentary record of the periods of literate history (Van Riper calls them "historical" archaeologists). Their interests, often local in character, centred on the Middle Ages, and faded out when traced back beyond Roman times into the shadowy pre-literate 'Celtic' period of 'monuments' such as Stonehenge. They

saw little reason to concern themselves with crude stone artefacts that were unconnected with any such sites, and might not even be of human workmanship. Among geologists, on the other hand, there had already been decades of debate about alleged mixtures of human and animal remains in the geologically recent deposits of the 'Drift' (or 'Post-Pliocene') period. In particular, the pitfalls of interpretation surrounding deposits in caves — in which most such mixtures had been found — were well recognized.

The evidence that precipitated the 'revolution' came first from the newly discovered Brixham cave in Devon: excavated with unprecedented care, this greatly strengthened the case for the contemporaneity of human beings and extinct mammals. That case was then swiftly clinched,



Mammoth among the men — new evidence found in the early nineteenth century greatly strengthened the case for the coexistence of man and extinct animals.

to the satisfaction of almost all the scientists involved, by finds confirming earlier claims that similar stone tools in the ancient gravels of the Somme were truly associated with the bones of extinct animals; for these gravels were regarded as much less problematic deposits than those in any cave. Van Riper's narrative of these events is the core of his book; it is a fascinating story that is sensitive to the tensions between the London élite and the local amateurs of Torquay and Abbeville, and to the social dynamics among the élite scientists themselves.

To round off, Van Riper traces the varied responses of wider groups in British society to the new conclusions of the

experts; and he shows how a new discipline of "prehistoric archaeology" was built on the territory of the "Old Stone Age" or Palaeolithic period. The contemporaneity of early human beings and extinct mammals, and the vastly enlarged scale of human antiquity that was its corollary (not uncontested), carried implications for man's place in nature that were as far-reaching as those of Darwin's *Origin*. Van Riper shows, however, that there was no simplistic conflict of 'science versus religion' on this issue, but a wide range of responses: for example, the new biblical criticism exemplified in the famous volume *Essays and Reviews* (published coincidentally at just this time) was quite as influential as any backwoods literalism. As for the rise of a newly defined discipline, Van Riper shows how

the proponents of "prehistoric archaeology" captured and transformed the older institutions so successfully that the study of pre-literate cultures, exploiting the methods of both geology and anthropology, became virtually synonymous with archaeology itself, banishing the earlier discipline to the sidelines.

Van Riper's story is almost entirely a British one, and he defends that perspective on the grounds that it reflects the cutting edge of the historical case for human antiquity. That is at least partly true. But a parallel debate of equal intensity was agitating Parisian savants at that time; after all, much of the ultimately decisive evidence was found on French soil. Van Riper acknowledges this in passing, for example when he deals with the celebrated case of the allegedly fraudulent Moulin-Quignon jaw (a forerunner of the Piltdown one, as it

were). More on the international dimension would have made his account still better. As it is, a tale of two cities is almost reduced to a tale of one; but it is nonetheless a tale finely told. Indeed, this book is a rarity: although it started life as a dissertation, it is beautifully written, in prose that is elegant yet simple and lucid, and it is a pleasure to read. Its highly visual subject matter deserved many more illustrations, but the book itself is produced to high standards. □

Martin Rudwick teaches the history of science in the Science Studies Program at the University of California, San Diego, La Jolla, California 92093-0104, USA.